Distributed Localization Algorithms in Indoor 802.15.4 Wireless Networks

N'Deye Amy Dieng

N° d'ordre : 2014telb0308

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En habilitation conjointe avec l'Université de Rennes 1

Ecole Doctorale - **MATISSE**

Algorithmes de localisation distribués en intérieur pour les réseaux sans fil avec la technologie IEEE 802.15.4

(Distributed localization algorithms in indoor 802.15.4 wireless networks)

Thèse de Doctorat

Mention : « Informatique »

Présentée par **N'deye Amy Dieng**

Département : Réseaux, Sécurité et Multimédia (RSM)

Laboratoire : IRISA

Directeur de thèse : Xavier Lagrange

Soutenue le 7 Janvier 2014

Jury :

M. Congduc Pham - Professeur, Université de Pau et des Pays de l'Adour (Rapporteur)
M. Hichem Snoussi - Professeur, Université de Technologie de Troyes (Rapporteur)
M. César Viho - Professeur, Laboratoire Irisa/Université Rennes 1 (Président et Examinateur)
M. Chérif Diallo - Maître de Conférences, Université Gaston Berger (UGB) (Examinateur)
M. Tayeb Ben Meriem - Professeur, Orange Lab Network Carrier/ R&D (Invité)
M. Maurice Charbit - Maître de Conférences, Télécom ParisTech (Invité)
M. Claude Chaudet - Maître de Conférences, Télécom ParisTech (Encadrant)
M. Laurent Toutain - Maître de Conférences, Télécom Bretagne (Encadrant)
M. Xavier Lagrange - Professeur, Télécom Bretagne (Directeur de thèse)

*** *To my beloved best friends, i.e. Mom and Dad, for their infinite love.*
Thanks for everything.
In memory of my grandparents and my friend Pape Amadou Diaw. ***

Acknowledgments, Remerciements, Jërëjëf

*** Lëp lo def ak diom, fouleu ak fayda, té gueum sunu borôm ak sa bop, ndam fèke la thi. ***

This thesis is the result of my passion for studying computer networks and Information Technology (IT). These years of research were a rewarding experience on both the scientific and human levels. I met researchers from various countries, who worked in different IT fields, and that was invaluable in my quest for knowledge. Shad Helmstetter said, "Apprendre, c'est déposer de l'or dans la banque de son esprit" (Learning is gold deposit in the bank of his mind). All of you have greatly contributed to my learning, thank you.

I would like to thank all the jury members for attending my PhD defense: Cesar Viho, professor at Université de Rennes 1, laboratoire IRISA: for accepting to be a member of the jury, Professors Congduc Pham, Université de Pau et des Pays de l'Adour and Hishem Snoussi, Université de Technologie de Troyes: for their efforts in reviewing my thesis and preparing a report of my scientific work, Dr. Cherif Diallo, Université Gaston Berger St louis: for accepting to attend to my PhD defense, many thanks to you all for your time and valuable comments, which contributed to the improvement of this thesis manuscript.

To my two thesis advisors, Dr. Claude Chaudet associate professor at Telecom ParisTech and Dr. Laurent Toutain associate professor at Telecom Bretagne: thank you for allowing me to pursue this research topic, for your guidance and for teaching me the fundamentals of scientific research. I was able to work autonomously on my thesis, and for that I am really grateful.

I am thankful for Orange Labs, which supported my research work, and for professor Tayeb Ben Meriem for trusting me as a PhD student.

To my supervisor, Xavier Lagrange professor at Telecom Bretagne: thank you for having me as a PhD student and overseeing my research work.

To Maurice Charbit professor at Telecom ParisTech, TSI laboratory Darrault with whom I worked countless times during these years of research: thank you, our discussions were very rewarding scientifically and humanly.

To Dr. Michel Grojnowski, TSI laboratory Telecom ParisTech Barrault: thank you for your good mood and your support. Our exchanges of ideas have guided my research.

I express my gratitude to Mr. Pouye, Professor Matta, and Professor Jong for their good advices, which inspired me throughout my studies.

Thank you to the RSM Team at Telecom Bretagne, including Labo4G and their scientific direction. I was welcomed at Telecom Bretagne with open arms and each visit at Rennes and Brest was a pleasure. Thank you for your good mood and for Breton crepes. I also thank all those who helped me during the experimentations with sensor nodes: Alexandru (LIP6), Benoit (Telecom Bretagne), Moussa and his advisor professor Roblin, Comelec department at Telecom ParisTech.

My warm thanks go to Gemma, PhD student at Telecom ParisTech, and her supervisor Dr. Pascal Bianchi. We had a productive collaboration and I enjoyed working with such a dynamic and sympathetic person. Thank you for reading the thesis manuscript.

Thank you all members of the INFRES department at Telecom ParisTech, Inria and Lincs laboratory, all PhD students and engineers at Inria, Lincs, Telecom ParisTech as well as Telecom Bretagne: Sameh, Giuseppe, Silvio, Stephano, Raphaelle, Chiara, Rim, Paulo, Claudio, Akhlem, Dorice, Ahmed, Mayssa,

Xavier, Sebastien, Matteo, Rosy, Gege, Ghida, Lucien, Sarah, Baptist, German. It was very pleasant and instructive to share these moments with you in a dynamic and international environment.

A big thank to all my friends and my lovely honey. Thanks to you, these years have been wonderful and especially rich in friendship.

Special thanks to all my family members who have always shown me their support during my research.

Finally, I express my deep gratitude to my mom and dad, my grandparents for their love, devotion and support. You have always trusted and supported me over these years. The values you have instilled in me always follow me and have helped me to climb the path of life with faith, humility, willingness and trust. Grandfather, from up there I know you are proud of me. Mom, I love you, you are the greatest blessing of my life, "mon jardin préféré dont la sagesse ne désemplit pas", *lou djig yaye diag, yaye borom bäg bi*. When *Sedar Senghor* sang the praises of women, it was meant for you. I do not have enough words to thank you, *Dieuredieuf yaye boye*.
"Rien ne vaut une douce maman avec un grand coeur." (Nothing is greater than a mother with a great heart). Léon Tolstoï.

Abstract

The Internet of Things is finally blooming through diverse applications, from home automation and monitoring to health tracking and quantified-self movement. Consumers deploy more and more low-rate and low-power connected devices that provide complex services. In this scenario, positioning these intelligent objects in their environment is necessary to provide geo-localized services, as well as to optimize the network operation. However, indoor positioning of devices using only their radio interface is still very imprecise.

Indoor wireless localization techniques often deduce from the Radio frequency (RF) signal attenuation the distances that separate a mobile node from a set of reference points called landmarks. The received signal strength indicator (RSSI), which reflects this attenuation, is known in the literature to be inaccurate and unreliable when it comes to distance estimation, due to the complexity of indoor radio propagation (shadowing, multi-path fading). However, it is the only metric that will certainly be available in small and inexpensive smart objects. In this thesis, we therefore seek algorithmic solutions to the following problem: is it possible to achieve a fair localization using only the RSSI readings provided by low-quality hardware? To this extent, we first study the behavior of the RSSI, as reported by real hardware like IEEE 802.15.4 sensor nodes, in several indoor environments with different sizes and configurations[1], including a large scale wireless sensor network. Such experimental results confirm that the relationship between RSSI and distance depends on many factors; even the battery pack attached to the devices increases attenuation.

In a second step, we demonstrate that the classical log-normal shadowing propagation model is not well adapted in indoor case, because of the RSSI values dispersion and its lack of obvious correlation with distance. We propose to correct the observed inconsistencies by developing algorithms to filter irrelevant samples. Such correction is performed by biasing the classical log-normal shadowing model to take into account the effects of multipath propagation. These heuristics significantly improved RSSI-based indoor localization accuracy results. We also introduce an RSSI-based positioning approach that uses a maximum likelihood estimator conjointly with a statistical model based on machine learning.

In a third step, we propose an accurate distributed and cooperative RSSI-based localization algorithm that refines the set of positions estimated by a wireless node. This algorithm is composed of two on-line steps: a local update of position's set based on stochastic gradient descent on each new RSSI measurement at each sensor node. Then an asynchronous communication step allowing each sensor node to merge their common local estimates and obtain the agreement of the refined estimated positions. Such consensus approach is based on both a distributed local gradient step and a pairwise gossip protocol. This enables each sensor node to refine its initial estimated position as well as to build a local map of itself and its neighboring nodes. The proposed algorithm is compared to multilateration, Multi Dimensional Scaling (i.e. MDS) with modern majorization problem and classical MDS. Simulation as well as experimental results obtained on real testbeds lead to a centimeter-level accuracy. Both landmarks and blind nodes communicate in the way that the data processing and computation are performed by each sensor node without any central computation point, tedious calibration or intervention from a human.

Index terms— Indoor Localization, Distributed Localization, Wireless Sensor Network, RSSI, Log-normal Shadowing, IEEE 802.15.4, Experiments, Simulations

[1] `http://nadieng.wordpress.com/`

Résumé

L'internet des objets se développe à travers diverses applications telles que la domotique, la surveillance à domicile, etc. Les consommateurs s'intéressent à ces applications dont les objets interagissent avec des dispositifs de plus en plus petits et connectés. La localisation est une information clé pour plusieurs services ainsi que pour l'optimisation du fonctionnement du réseau. En environnement intérieur ou confiné, elle a fait l'objet de nombreuses études. Cependant, l'obtention d'une bonne précision de localisation demeure une question difficile, non résolue.

Cette thèse étudie le problème de la localisation en environnement intérieur appliqué aux réseaux sans fil avec l'utilisation unique de l'atténuation du signal. L'atténuation est mesurée par l'indicateur de l'intensité du signal reçu (RSSI). Le RSSI est connu dans la littérature comme étant imprécis et peu fiable en ce qui concerne l'estimation de la distance, du fait de la complexité de la propagation radio en intérieur : il s'agit des multiples trajets, le shadowing, le fading. Cependant, il est la seule métrique directement mesurable par les petits objets communicants et intelligents. Dans nos travaux, nous avons amélioré la précision des mesures du RSSI pour les rendre applicables à l'environnement interne afin d'obtenir une meilleure localisation. Nous nous sommes également intéressés à l'implémentation et au déploiement de solutions algorithmiques relatifs au problème suivant : est-il possible d'obtenir une meilleure précision de la localisation en utilisant uniquement les mesures de RSSI provenant des nœuds capteurs sans fil IEEE 802.15.4 ?

Dans cette perspective, nous avons d'abord étudié le comportement du RSSI dans plusieurs environnements intérieurs de différentes tailles et selon plusieurs configurations[2], y compris un réseau de capteurs sans fil à grande échelle (SensLAB). Pour expliquer les résultats des mesures, nous avons caractérisé les objets communicants que nous utilisons, les nœuds capteurs Moteiv TMote Sky, par une série d'expériences en chambre anéchoïque. Les résultats expérimentaux confirment que la relation entre le RSSI et la distance dépend de nombreux facteurs même si la batterie intégrée à chaque nœud capteur produit une atténuation. Ensuite, nous avons démontré que le modèle de propagation log-normal shadowing n'est pas adapté en intérieur, en raison de la dispersion des valeurs de RSSI et du fait que celles-ci ne sont pas toujours dépendantes de la distance. Ces valeurs devraient être considérées séparément en fonction de l'emplacement de chaque nœud capteur émetteur. Nous avons proposé des heuristiques pour corriger ces incohérences observées à savoir les effets de la propagation par trajets multiples et les valeurs aberrantes. Nos résultats expérimentaux ont confirmé que nos algorithmes améliorent significativement la précision de localisation en intérieur avec l'utilisation unique du RSSI.

Enfin, nous avons étudié et proposé un algorithme de localisation distribué, précis et coopératif qui passe à l'échelle et peu consommateur en termes de temps de calcul. Cet algorithme d'approximation stochastique utilise la technique du RSSI tout en respectant les caractéristiques de l'informatique embarquée des réseaux de capteurs sans fil. Il affine l'ensemble des positions estimées par un nœud capteur sans fil. Notre approche a été comparée à d'autres algorithmes distribués de l'état de l'art. Les résultats issus des simulations et des expériences en environnements internes réels ont révélé une meilleure précision de la localisation de notre algorithme distribué. L'erreur de localisation est de l'ordre du centimètre sans aucun nœud ou unité centrale de traitement, ni de calibration fastidieuse ni d'intervention humaine.

Mots clés—Localisation environnement interne, Localisation distribuée, Réseau de capteur sans fil, RSSI, Log-normal Shadowing, IEEE 802.15.4, Expérimentation, Simulations

[2] `http://nadieng.wordpress.com/`

Contents

List of Acronyms

AOA Angle Of Arrival

AWGN Additive White Gaussian Noise

AP Access Point

APIT Approximated Point-In-Triangulation

CDF Cumulative Distribution Function

EM wave Electromagnetic wave

EWMA Exponentially Weighted Moving Average

GMM Gaussian Mixture Models

GSM Global System for Mobile communication

IR Infrared

IP Internet Protocol

KBps Kilobytes (i.e. 1024 bytes) per second

LNNE Localization-based Neural Network Ensembles

LNSM Log-Normal Shadowing Model

LNSM-DV Log-Normal Shadowing Model with Dynamic Variance

LOS Line of Sight

LR WPAN Low-Rate Wireless Personal Area Networks

m Meter

MAC Medium Access Control

MIMO Multiple Input Multiple Output

MLE Maximum Likelihood Estimator

NLOS Non-Line of Sight

NME Normalized Mean Error

PB-ROCRSSI Power-Based ROCRSSI

PCB Printed circuit board

PDF Probability Distribution Function

RAMMUS RAdio Mobile Multi Standards

RMSE Root Mean Squared Error

RF Radio Frequency

RFID Radio Frequency Identification

ROCRSSI Ring Overlapping Based on Comparison of Received Signal Strength Indicator

RTOF Roundtrip Time Of Flight

SMS Short Message Service

SNR Signal-to-Noise Ratio

ToA Time of Arrival

TDoA Time Difference of Arrival

UMTS Universal Mobile Telecommunications System

UWB Ultra Wide Band

WLAN Wireless Local Area Network

WSN Wireless Sensor Networks

Chapter 1

Introduction

Contents

1.1 Prologue and motivation

****La science est ce que nous comprenons suffisamment bien pour l'expliquer. On fait la science avec des faits, comme on fait une maison avec des pierres : mais une accumulation de faits n'est pas plus une science qu'un tas de pierres n'est une maison.* (—Henri Poincaré—) ***

1.1.1 Background of the study

Progress in electronic and signal processing allows the design of low power, low cost wireless devices leading to the concept of Internet of Things (IoT). IoT, also called internet of "Intelligent Things", is the ability of physical objects with an identified IP address (and associated with embedded technology) to communicate over a network of "everything"[1] [2]. Examples of physical objects include medical devices, electric meters and heart monitors. IoT is a sort of advanced connectivities on a global scale to encompass every aspect of our life in order to render it much easier and safer without the necessity of human-to-human or human-to-computer interaction.

"Things/devices" connected in an internet structure are becoming increasingly diverse. Beyond various connected accessories, phones, computers, there are lots of different concepts, including everything about the smart home, household appliances and home automation. Such connected "things or devices" in an internet-like structure, involve a complete data infrastructure in the back end like machine-to-machine, machine-to-infrastructure, machine-to-environment and human-to-machine.

Some applications which may be involved in the IoT require a precise localization. For instance, applications like geographical tagging of measurements, network operation (e.g. geographic routing), personal health monitoring, objects or persons tracking, intelligent homes and cities as well as environment study and preservation. In the smart home context, there are still a lot of applications more or less convincing, but all share a number of characteristics: they work

[1] `http://www.gartner.com/newsroom/id/2564916`
[2] `http://iotinternetofthingsconference.com/`

on devices with low capacity, which are in part mobile and deal with geo-localized services (i.e LBS). Note that users are not necessarily experts and we do not want them to configure many things on their own hand.

This thesis focuses on the problem of indoor localization with the assumption that devices only have a radio interface. There are several reasons for a mobile device to acquire its location automatically in an indoor environment. Indoor localization cannot use GPS signals efficiently, due to signal obstruction: GPS signals do not penetrate physical indoor concrete structures. Thus, before using localization information in the Internet of Things applications, several issues need to be addressed.

First, the well-known RF technology, available in most wireless devices, is difficult to characterize in an obstructed environment where walls and concrete structures have a strong impact on the wireless radio signal.

Second, the mobility context, the indoor setup and layout, make localization challenging. In fact, the mobility of devices rules out pre-assignation of localization information for several terminals and requires the design of an online location estimation method. However, the unpredictable nature of the signal propagation in indoor condition prevents from obtaining accurate location information. Indoor localization has been the subject of many studies and still remains a very difficult issue. With the instability of the indoor environment in which signal may be blocked, dispersed or reflected, most RSSI (Received Signal Strength Indicator)-based indoor localization algorithms accuracies are roughly about one or more meters. In applications like smart home, such algorithms are either very complex for their use of calibration or initial training phase [2, 3, 4] or not accurate enough [5, 6, 7]. These algorithms are neither suitable for the surveillance of elderly people in nursing homes, nor for tracking users in a mall.

Third, indoor localization using wireless sensor networks (WSN) is not trivial as it should meet some requirements. A wireless sensor network (WSN) is composed of devices, small in size, capable of multifunctional sensing, treating, processing data and communicating in short distances. These tiny devices usually called sensor nodes are also limited in terms of processing, storage capability and they often have to keep energy consumption low. Sensor nodes in a WSN are therefore built upon unique and fundamental properties. This is particularly, as mentioned, the very limited processing capabilities and energy of individual nodes.

A wireless sensor network [8] presented in Figure 1.1, is also defined as a result of deployment of several tiny sensor devices (i.e. interconnected micro-components) in environments whose properties are unknown. Then, each communicating micro component is equipped with a capture unit to collect data from their environment; a low memory as well as a battery pack to power various components and an emission/reception radio module for data communication. Collected data can be related to temperature, the humidity ratio and the speed. Collected data are analyzed with a built-in computing unit in sensor node. For systems such as Moteiv TMote Sky nodes, the cost and limitations of the hardware on sensing nodes must prevent the use of costly algorithms in terms of computation, memory usage and energy consumption. Sensor nodes in consequence require
using collaboration among nodes while minimizing communication. Therefore when deploying an algorithm or communication protocols for sensor nodes, designers must pay attention to energy consumption. The above properties of sensor nodes characterize them as low power wireless devices with real embedded systems and therefore do not facilitate the design of application software.

Besides aforesaid, wireless sensor network is fundamental in many applications for data collection

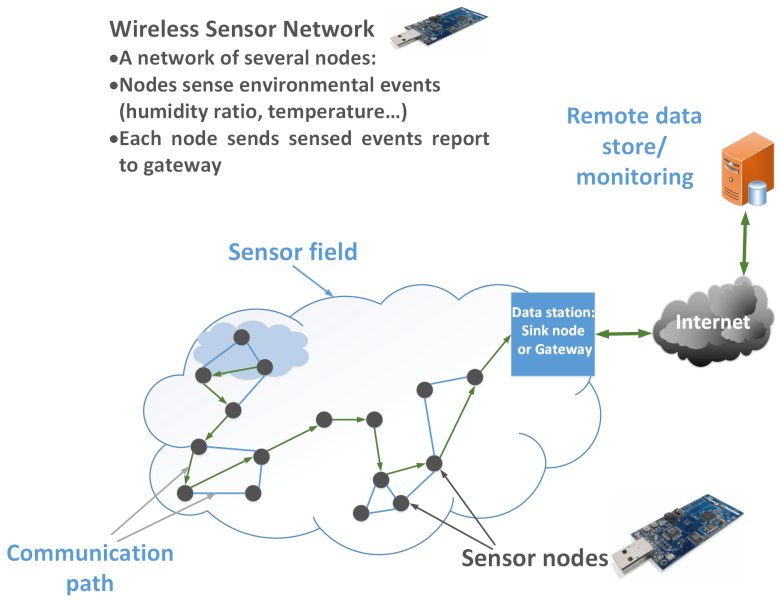


Figure 1.1: An example of Wireless Sensor Network system architecture

and data processing. Collected data is used in localization algorithms, thus enabling localization in sensor network can automatize and bring new features for many applications in many areas. Examples of tiny sensor devices, such as Moteiv TMote Sky nodes, can be envisioned in the IoT not only for application control, monitoring and surveillance but also for location-based services. Location-based services with tiny devices can also be used in health care domain to enable patient location and monitoring at medical-home or in the hospital. For example, transmission of diseases can be studied through automatic measurements of blood pressure/blood glucose/heart rate of a patient and transmission by health sensor nodes. Wireless sensor network is also useful in disaster prevention, in environmental monitoring, in fire and rescue system to monitor warehouse/building fire and direct firefighters appropriately.

Depending upon the type of application involved, the number of sensor nodes may relatively vary. For example within a home it is between ten to fifty or more. In addition, most of the sensor nodes should be in static positions or attached to humans and objects in movement. For each of these examples, the acquired data is meaningless if no location information is associated. For such reasons, localization in WSN is needed for a better understanding of the corresponding environment.

Localization in indoor environment can bring a new framework of ideas like triggering services in an intelligent home for easy home management. Home based services can be associated to a position of the occupant using a WSN measurements as shown in Figure 1.2.

However, an understanding of disrupting factors and how they impact radio wave propagation in indoor and the design of an accurate, adaptive and easy to use indoor localization are still a challenging research problem. To address such problem, we have designed robust indoor radio frequency-based localization algorithms. Proposed algorithms are tested on wireless sensor nodes and are adaptive to environmental dynamics. Algorithms do not require any initial training phase or site-survey. Those indoor localization algorithms are scalable to any indoor environment (large or small). They are also built to capture interference and propagation disruption and hence measurement errors.

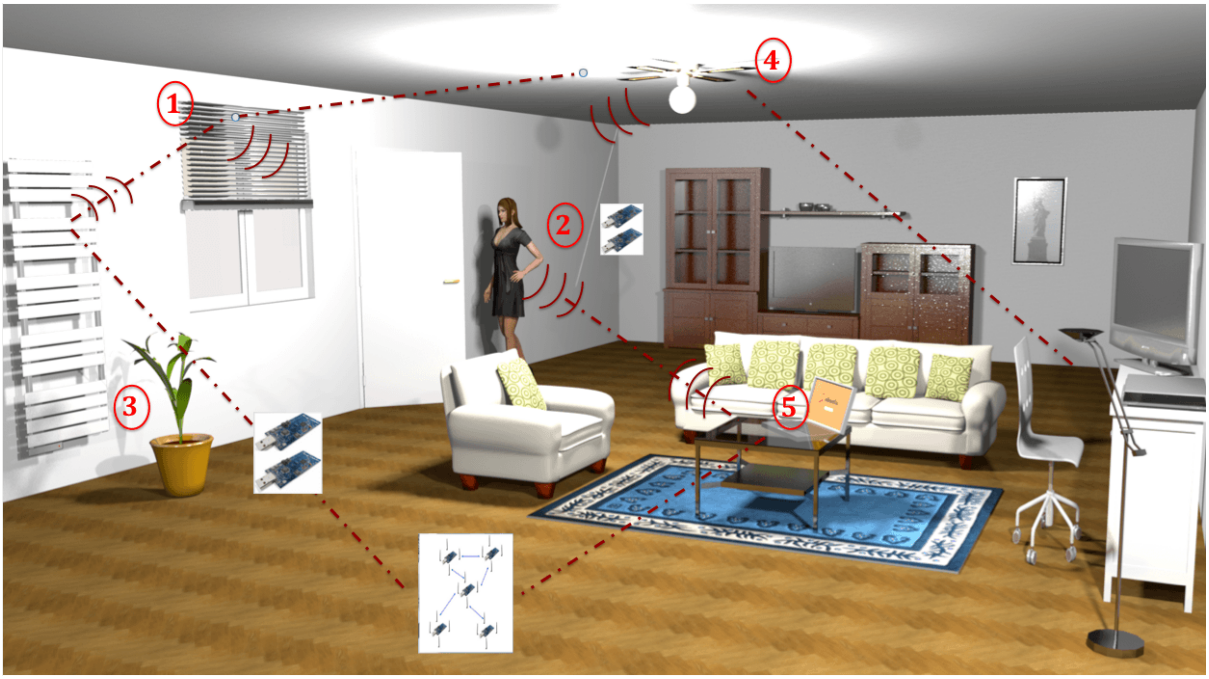


Figure 1.2: An example of application in smart home: all devices in a home can be automatically and simultaneously controlled depending on the position of the occupant. This involves an interaction between the positioning system and the remote control services. When the occupant leaves or arrives at home, shutters can automatically be closed or opened (1), security alarm can automatically be deactivated or activated (2), temperature, humidity sensors and radiators can be automatically regulated to maintain ideal temperature (3); electricity, e.g. lighting is also easily controlled without action (4). At the end, notification are directly sent to the occupant for any feedback (5).

1.1.2 Problem Statement

Localization is a set of raw or abstract ideas about the location of something or someone. It determines the relative or absolute position of a target in a 2D or 3D system. The location estimation is referred to as positioning. The localization accuracy is defined as the distance error between the estimated position and the real position. Positioning a device in a coordinated system is possible via several techniques. Such techniques can be classified based on the nature of the signal and the measurement method used to estimate the location.

- Non-radio frequency ranging devices such as laser telemeters are too expensive for embedded devices and require a clear line of sight between the mobile and each landmark. Indeed, the line of sight constraint rules out all light-based ranging techniques such as infrared transmitters [9], [10], [11]. Ultrasound [11] is fairly accurate, but also requires a dedicated hardware (e.g. a more expensive transducers) that is usually not present on consumer devices. Ultrasound uses ultrasonic waves to measure the distance between an ultrasonic emitter and receiver. Most algorithms that use ultrasonic systems are mostly hybrid since the propagation speed of the sound is too slow compared to a radio frequency signal. For instance, Priyantha *et al.* combine ultrasound and RF signals in *cricket* location system [12]. Instead, radio frequency-based techniques, do not require any specific electronics. RF interfaces are already used for packet communications. All these mentioned technologies use a common feature as a wireless device (an emitter or a receiver node/tag) needs to be carried by a person or an object.

- As said before, localization-based on radio frequency technology is cost effective. A well known used approach is the *proximity* technique [13]. It is easy to measure as the simple use of the radio connectivity is enough to assess whether a mobile device is within the transmission range to a reference device. For instance, a reference device can be a Wi-Fi access point to which a mobile node or a user's smartphone is linked. However, *proximity* technique has an accuracy limited to the one of the reference equipment and equal to one transmission range. Note that if multiple reference equipments detect a mobile device, this device position is the position of the reference equipment receiving the strongest signal. Hence, this *proximity*-based method gives therefore coarse location. Resulted location with *proximity*-based method is not enough accurate to be applied for instance in smart home context in which a fine grained localization could be needed in most cases. *Proximity*-based method may also require association with an identification system, if it is not included in the method of proximity detection. Examples of systems based on this technique are infrared (IR) [9] and radio frequency identification (RFID). There are two types of *proximity* approaches for localization: detection of physical contact (pressure sensors, smart floor) and monitoring of wireless access points. The works of Partridge *et al.* [14] show a mechanism based on this principle and named the "fast intrabody signaling", which is a communication mechanism that transmits data electrically through human body. The identification phase therefore requires human contact enabling communication between objects touched by the user in order to exchange identification information through the same communication channel.

- Meanwhile using the empirical model may be more accurate but requires a site survey. Commonly named as *scene analysis*, it matches the radio propagation environment characteristics to a database [15, 16]. In other words, *scene analysis* examines the differences between two successive scenes (e.g. a place or a view) to position any target that has been observed in a scene. For instance, in [16], *Scene analysis* is used to compute the position

of the camera that has observed the horizon using the extracted samples from the horizon. Thus, scene analysis requires to perform a pre-calibration of the environment characteristics. This training phase is required by this technique to perform prior measurements in the concerned area. Those measurements are then maintained in a database and subsequently used during the localization resolution phase. Therefore, a localization accuracy using scene analysis technique is limited to the measurement environment. Position can also be determined by RF fingerprinting [2, 17, 18]. In an offline phase, RSSI measurements are collected at several positions in the concerned area from data sent by known reference nodes (e.g. Wi-Fi access points or 802.15.4 nodes). These measurements named training set are then recorded in a database. The data collection step is time consuming.

During the online phase, the position of a mobile node is determined using a pattern matching mechanism. This latter matches the current RSSI measurements to those from a database using deterministic or probabilistic approaches. In [19], the deterministic approach of K nearest neighbors is used to localize a mobile user from three Wi-Fi access points in a floor building with dimensions 43.5 m by 22.5 m. Indeed, average signal strength from K nearest neighbors are matched to the RSSI measured during the online phase. The authors demonstrated that $K = 3$ nearest neighbors result in a best localization accuracy ranging from 2 to 3 meters. This obtained localization accuracy corresponds to the size of an office room in the tested building.

- Instead of considering the radio proximity or *environment features*, localization systems can also be classified as *distance* or *angle-based*. Distance-based approaches can be performed through the measurement of the propagation time or the attenuation experienced by a receiver. Attenuation is the power loss on the communication path between an emitter and a receiver. Distance-based (i.e. related to time-based, RSSI-based) and angle-based measures are feed into localization algorithms like trilateration to determine the positions of the mobile nodes as illustrated in Figure 1.3. Trilateration or triangulation techniques [1, 20], use distance or angle measurements from at least three emitting wireless devices when considering $2D$ position coordinates. Such localization techniques require a precise and predictable relationship between the distance or the angle and the measured parameter (i.e. time, RSSI, angle).

Using the wireless interface to measure distances between communicating objects may not require additional hardware. However, the only easily measurable parameter related to distance is the attenuation, which has low accuracy. The time of flight of the radio signal yields a better accuracy but is more difficult to measure, as it requires a fine time synchronization between communicating nodes. Angles measurements can replace distances in the geometric model but require the reception antenna to be able to precisely identify the direction of the direct path. Therefore, signal attenuation, usually available through the RSSI, appears as the only economically feasible ranging solution for the IoT.

However, the RSSI measures are far from constant in an indoor environment due to the wireless radio channel instability. The radio channel being the physical support for the transmission of information in wireless communication, is unsteady, random and lossy and thus limits the performance of wireless communication systems. Phenomena such as shadowing, Non-Line of Sight (i.e. NLOS) or situations of non-visibility, multipath fading, affect signal propagation. Such wireless characteristics change with furniture, the presence of people moving inside the environment and altering at the same time the narrowband channels like the one used by the IEEE 802.15.4 norm. These effects make ranging based on RSSI technique inaccurate for some specific needs such as artwork surveillance in museums. Consequently, precise RSSI-based indoor

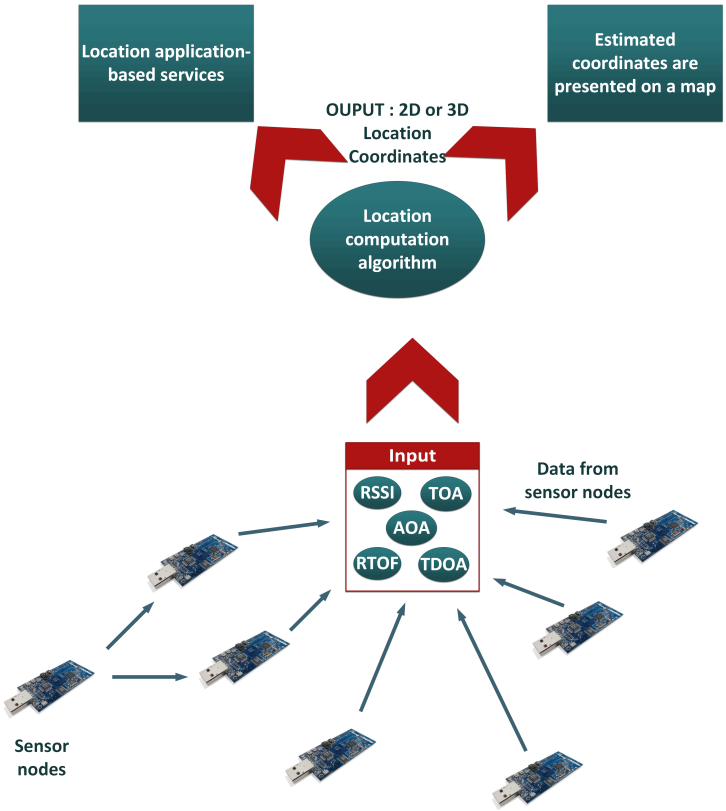


Figure 1.3: Localization algorithm process

localization algorithms require a characterization of the environment and of the radio channel as accurate as possible.

The radio wave propagation varies according to environmental parameters [21]. As a result, the shadowing effect models the wireless radio signal by considering several potential obstacles. Given a distance that separates a wireless emitter to a wireless receiver, different environment cluttering may disturb the received signal strength [22, 23].
Shadowing occurs with the presence of an obstacle obstructing the radio communication path. Such obstacle can disrupt the direct path between an emitter and a radio receiver giving some areas of weak received signal strength. Hence, the received signal power fluctuates as the direct path (line-of-sight) can be received with multiple other indirect paths. Direct path (i.e. line-of-sight) modeled in Figure 2.1 is viewed as a straight propagation of radio waves without any obstacles between the emitter and receiver antennas. Fading is obstruction that emanates from a deviation of the attenuation due to the presence of multiple paths during the propagation of the radio wave. The presence of one noticeable predominant path is called line of sight or Direct line of sight path. In opposite, the absence of such dominant path results in non line of sight paths. Consequently several efforts have been made in recent years to better characterize the corresponding statistical model to measurements errors in order to improve RSSI-based indoor localization accuracy. Thus, indoor positioning using RSSI metric is still a challenging research topic.

The first objective of this PhD thesis is to understand and to better characterize the indoor channel behavior. As a starting point, to circumvent calibration phases, we use a radio propagation model. Such formal model does not implicate explicit measurement of radio frequency signal strength at several locations but uses a mathematical model to characterize the signal

propagation [19]. We select a propagation model in order to deduce distance from the attenuation experienced by radio packets measured by each sensor node. The selected radio propagation model should represent a good compromise between accuracy and complexity. Indeed, the localization algorithm has to be performed on embedded devices and hence should not require too much computation or data exchange. The Log-Normal Shadowing Model (i.e. LNSM) [24, 25], often used to model indoor propagation, seems to be relevant in this situation. The input parameters to this propagation model describe the different obstacles that may be present in the communication path between the radio emitter and receiver. An observation model is called log-normal shadowing if the attenuation or signal loss, expressed in deciBel (i.e. dB), has a Gaussian distribution.

To conclude, we then designed accurate RSSI-based indoor positioning algorithms that consider and treat above mentioned wireless signal propagation perturbations. Evaluated with sensors data, our proposed RSSI-based localization algorithms can be applied in any indoor environment (office, building, house) with no complementary pre-deployment efforts.

1.2 Contribution of this PhD thesis

In this thesis, we:

- introduce the background material/knowledge on radio signal propagation model by understanding the behavior of the indoor radio propagation model through the use of the Received Signal Strength Indicator (RSSI) metric. We base our experiments on 802.15.4 wireless materials that we believe to be representative of low power and low coast wireless devices.
- investigate the implication of the log-normal shadowing model in indoor propagation: by clarifying its relationship with the measured RSSI in an indoor environment with presence of furniture, obstacles and obstructing objets.
- provide a set of experimental results based on extensive measurements campaigns in a variety of indoor environments. Such experimental results are needed in order to validate the use of the RSSI as a unique metric for localization in realistic indoor scenarios.
- identify effective statistical tools based on maximum likelihood optimization, machine learning as well as distributed stochastic approximation approach to improve location accuracy of a mobile wireless sensor node in the indoor environment.
- propose efficient and non-calibration RSSI-based algorithms for positioning in real indoor environment. For these algorithms, we provide a detailed comparisons in both: realistic indoor testbeds as well as a large scale wireless sensor network.
- evaluate and show by multiple comparisons the performance of proposed algorithms and the conditions for possible future improvement.

In our research work, we propose and evaluate a family of one-hop RSSI-based localization algorithms. Our RSSI-based localization algorithms enhance indoor positioning of wireless devices without consideration of an offline pre-calibration/measurement phase, neither a human intervention nor a maintenance of a signature database. We use RSSI measurements from data packets coming from multiple sources called *landmarks*. Given landmarks coordinates and RSSI

readings, a mobile node evaluates its own position by measuring its distance from each landmark and using a localization algorithm like trilateration or multilateration. If landmarks are not aware of their own position, the computed position can still be deducted but relative to the local scene. The difficulty in this process resides in the correct estimation of the distances, also called ranging to the landmarks. Notice that the data processing as well as the estimation and computation of a blind mobile node position should be performed by a blind mobile node itself to allow privacy, instead of using a dedicated server via a base station or a coordinator or a sink node. For example, a base station centralizes the localization process as performed in [26]. As a starting point, we studied the behavior of the RSSI by carrying out experimental studies through extensive measurement campaigns in several indoor environments and in large scale sensor networks (available from the SensLAB platform).[3]. Such uncontrolled scenario shows that if some sources are coherent, some are clearly incoherent with the others. This phenomenon can be explained by the radio propagation (e.g. multipath) or by poor hardware and when excluding these incoherent data sources from the data set should lead to a better accuracy.

We then propose a dynamic method to alleviate the effect of multipath by identifying abnormal landmarks and reduce their effects, replacing their measurements by a constant bias. We extend the algorithm to multiple abnormal sensors and propose a heuristic to select the eluded sources quickly. We also evaluate the use of a *two-modes* Gaussian Mixture Model (GMM) to reduce the effect of outlier measurements. GMM [27] is a statistical machine learning approach which models an unknown random variable as the sum of multiple Gaussian random variables. *Two-modes* GMM decomposes the pattern distribution as a big percentage of normal patterns and a small proportion of outliers. Note that it does not need to know the exact number of outliers nor the outlying patterns. The $2D$ coordinates of a blind mobile sensor node is then computed in a realistic indoor environment by applying a maximum likelihood estimator. Such localization algorithm, named GMM-MLE, is then followed by an accurate on-line distributed stochastic approximation approach that minimizes the mean square error of the estimated distance derived from the noisy RSSI data. Recall that each RSSI data is measured from received data packets sent by neighboring sensor nodes.
The proposed distributed iterative algorithm aims to solve the localization problem of any blind nodes while taking into account indoor disrupting factors by treating the interference induced by various obstacles that compose the indoor environment: concrete walls, plywood, roofs, cabinets, fixtures, mobile person/user, geometry of the environment. Note that, known cooperative-based localization techniques estimate unknown nodes positions based on newly estimated positions of other nodes in the network in addition to those from known landmarks as well as connectivity information [28, 29, 30, 31]. However, the main problem affecting localization accuracy is the underlying model used to infer distance and the environment propagation; and such well-known methods do not focus on these problems. Consequently, the applied approaches limit the performance obtained after the refinement phase as they may ensure an appropriate initialization for the minimization criteria which could drive to the global optimum. Thus, if a high ranging error is acquired during the initial phase, such errors may affect the core algorithm, even with the use of a refinement phase.

With our proposed Distributed Stochastic Approximation Localization algorithm named DSAL, a set of blind nodes estimate their positions and those of their neighbors in an iterative and a cooperative way through two on-line phases: a first phase involving a local update of position's set based on stochastic gradient descent of [32] depending on each new RSSI measurement at each sensor node. A second phase concerns an asynchronous communication step [33] which allows each sensor node to build a local map of itself and its neighboring nodes. This phase

[3] `www.senslab.com`

is an exchange of information between two randomly selected nodes to reach a consensus on the set of estimated positions. Such consensus technique enables nodes to merge their common local estimates and obtains the agreement of the refined estimated positions. The refinement phase is completed when all nodes gets all updates from their neighbors. In other words, the algorithm converges if all nodes have perform refinement. Note that the set of landmarks do not need to refine theirs positions as they are known in advance. Numerical results are illustrated from simulated data as well as real data acquired in a large scale wireless sensor nodes named SensLAB Rennes platform. The RSSI-based DSAL cooperative algorithm has been compared through simulation and real data with multilateration and both classical and modern Multi-Dimensional Scaling (MDS) techniques.

We additionally emphasize that most RSSI-based indoor localization algorithms are mainly based on simulations. Our proposed algorithms are evaluated under various conditions in realistic indoor testbeds as well as in SensLAB large scale ad-hoc sensor networks with 250 nodes for scalability testing. Simulated algorithms do not guarantee the same kind of result when applied in a dynamic indoor environment. In all of our experimental localization scenarios, sensors are placed on grids or randomly distributed across the environment to meet the real deployment of a sensor network.

1.3 Thesis outline and methodology

As a first contribution we have performed extensive measurements to study and further analyze the way the measured RSSI values behave in different indoor environments (i.e. classrooms and offices building) and in large scale sensor networks (available from the SensLAB platform).[4].

Furthermore, localization algorithms using electronics devices must include hardware limitation in addition to physical constraints in order to incorporate the dynamic of real-indoor environments. Thus, measurements from different indoor environments are exploited such as in an anechoic chamber to characterize the radiation pattern of the sensor nodes used during our research work. In fact the characteristics of real antennas may deviate to some extent from those calculated analytically on sensor node datasheet.

Hence, having reliable antenna radiation patterns lead to the most efficient use of the radio spectrum and a better understanding of the wireless node characteristics. Consequently, we found that the used wireless node (i.e. *Tmote Sky* mote) radiation pattern is not fully omnidirectional.

Such measurements permit to achieve the following two goals:

- determine the relationship between measured RSSI and distance focused on signal strength variation in both time and space dimensions (i.e. in a mobile node context).
- understand and study the impact/influence of different indoor propagating environments on the RSSI behavior. The purpose is to validate the log-normal shadowing model for the use in any indoor environment under certain conditions. In fact, the log-normal shadowing model is the widely used propagation model in indoor environment and it is usually assumed to perform well in indoor and outdoor environment [24]. We unveil from measurement results that: depending on its environmental location, a receiving node may measure different behaviors on the RSSI values even if the distance to the emitting node remains the

[4] `www.senslab.com`

same. Therefore we have validated the idea to be considered that the log-normal shadowing model parameters depend on the environmental locations of the wireless nodes. Such result suggests the need to consider a modification of the log-normal shadowing model. We also provided tools and guides to follow in order to implement a more accurate and easy to use RSSI-based localization algorithm being suitable for any indoor environment.

It has come to our attention that, in the state of the art, already proposed indoor localization algorithms need to consider a large number of RSSI measurements, a selective mode of the emitting power, a less interfered and a high controlled indoor environments.

These measurements are devoted in Chapter 3 of this thesis. Our study is initiated by analyzing previous research papers. We focus on location estimation methods using received signal strength indicator (RSS) of RF hardware in 802.15.4 low-rate wireless personal area networks. We do not consider centralized localization approaches which need a central unit: the blind mobile node estimates itself its position.

In Chapters 4 and 5 of this thesis we focus on enhancing the precision and accuracy of one-hop and distributed indoor positioning algorithms based on the use of the RSSI metric. Note that, due to complex indoor propagation, accuracy still remains an important issue when designing indoor localization algorithms. As we mentioned before, several proposed RSSI-based indoor localization algorithms suffer on providing a reusable and accurate algorithm for any type of indoor environment. In fact, regardless of the technique used, the accuracy of localization comes at a price: as radio propagation is greatly influenced by the environment, knowledge of this environment is required from a calibration phase or through a specific deployment. This is a limiting factor when it comes to conceive generic and low-cost services.

For this reason, we use statistical approaches to solve this accuracy issue while avoiding the use of calibration. Moreover, we introduce a novel cross-online localization approach. This proposed online approach first estimates the propagation model parameters using a set of measurements coming from a subset of positions. Second, these parameters are used by the positioning algorithm. This cross-online approach provides a reliable indoor localization mechanism based on RSSI metric. In fact, a mobile target node can accurately position itself in any indoor environment while at the same time exploiting the site specific information of that environment. The propagation model parameters are in the meantime learned when performing positioning.

Proposed algorithms are evaluated and compared with known localization algorithms in realistic indoor environments. The aim of such investigation is really needed to quantify the level of accuracy provided by our proposed RSSI-based localization algorithms. The proposed algorithms meet some of their design goals, as for instance, they can be used with efficiency and acceptable accuracy in any wireless indoor environment without the need of any central station.

We discuss in the same Chapters 4 and 5 the main results achieved in this thesis. Finally, we draw in Chapter 6 the future direction, evolutions and conclusion of this study. Table 1.1 presents the overview of our research plan.

Table 1.1: Summary of the thesis organization

Chapter	Methodology	Comments
Chapter 1	Introduction and problem statement	Analyze of existing localization algorithms, current low power wireless technologies
Chapter 2	Related work	State-of-the-art of radio propagation models; related work on localization algorithms
Chapter 3	Measurements campaign in several real testbeds and in remote SensLAB large scale wireless sensor node platform	Sensitivity analysis of RSSI behavior under dynamic indoor environments, reverse engineering approach. Such measurements campaigns results are used in Chapters 4 and 5
Chapter 4	Study of most representative RSSI-based one hop indoor localization; proposition of new localization mechanisms, algorithms performance estimation, comparison with most used RSSI-based localization algorithms	Experiment, simulation, discussion
Chapter 5	Proposition of a new distributed localization mechanisms, comparison of our localization algorithm performance to known distributed state-of-the art localization algorithms	Experiment, simulation, discussion
Chapter 6	Conclusion	Conclusion and suggestions for further research

Chapter 2

Related work

Contents

In this chapter, we first present the methods of radio wave propagation. We then describe the most relevant radio frequency-based localization algorithms and the most used radio propagation models.

2.1 Indoor channel characterization: radio propagation models

The design of a wireless communication system with the use of the RSSI involves a selection of a propagation model for radio waves. Such propagation model depends on dispersion phenomena of wave propagation, on the selected frequency and on both the emitter/receiver antenna characteristics. Dispersion in time, direction and frequency are important aspects of radio communication, especially in indoor. Radio propagation techniques use mathematical models to describe these characteristics of radio channel. At lower frequencies (below 30 MHz) radio signals travel along ground waves while at higher frequencies (above 30 MHz) and in lower levels of the atmosphere like indoor environments, an emitted signal propagates along a line-of-sight (LOS) [34]. Thus, in a wireless communication system, especially in indoor, a clear LOS is needed and any obstruction between the emitter/receiver may block the signal.

Since the wireless medium is shared among different resources in a regulated manner (through MAC protocols) and a frequency band is attributed to each standard, there is no need to acquire a license to operate on the 2.4 GhZ ISM band. Different wireless standards may share the same frequency band, which can result in interference and channel overlapping [35]. For example, the 2.4 GhZ band is shared among bluetooth, Wi-Fi (i.e 802.11b/g/n), microwave and ZigBee/IEEE 802.15.4 devices. Sharing the frequency band can be particularly unfavorable for ZigBee/IEEE 802.15.4 networks [36]. Hence, the spectrum in this frequency band is highly busy, crowded and causes an increasing path loss and a limited range [37]. Path loss is the estimation of the signal attenuation or all lossy effects based only on the separation distance between an emitter and

a receiver. The path loss of the radio channel can also be defined as the ratio of the emitted power to the received power in decibel (dB) as:

$$\mathrm{PL}(dB) = 10\log_{10}(\frac{\mathcal{P}_e}{\mathcal{P}_r})(dB), \tag{2.1}$$

DeciBel, shortened as dB, is a measurement unit that refers to ratios and gains of an antenna on a log scale. A decibel is equal to $\frac{1}{10}$ of a Bel. Indoor narrow band communication system is profoundly affected by power dissipation and obstruction on the propagation channel. Consequently, propagation models are used to predict signal variation and provide a reliable estimate of the average signal strength in coverage area. In dynamic/mobile scenarios, statistical models are used and they provide statistical parameters to predict the propagation characteristics of the coverage area [37]. Maxwell's equations bound conditions that express the physical characteristics of obstructing objects [37]. Solving Maxwell's equations provide details on the electromagnetic wave which propagates through either reflecting or scattering or diffracting environment. However, this requires solving complex mathematics expressions.

Reflection, represented in Figure 2.1, happens when the propagating electromagnetic wave named EM go through a smoothed object or surface with dimension superior to the EM wavelength. Reflecting objects/surfaces can be furniture, walls, surface of the ground. The coefficient of reflection depends on the properties of the reflecting object/surface (frequency of the wave, angle of incidence, etc). Reflection causes the wave to be separated into two parts, one bypassing the object (e.g. wall, floors) and the other one reflected back into the device who originally sends the signal.
Scattering is caused by large rough surfaces, small objects with small dimensions compared to the wavelength of the electromagnetic wave. Scattering can also result from other irregularities such as stairs within buildings (for indoor environment), rough surfaces, lamp post (outdoor environment), foliage. The last but not least, diffraction propagation effect takes place when an electromagnetic wave encounters an obstacle (diffracting object) of wavelength comparable to the wavelength size. Diffraction is the result of the interference of scattered waves at each point of the diffracting object. Most often, diffraction results from the contact between an electromagnetic wave and the edger of an obstacle.
Diffraction can be observed with light, but also with sound, X-rays and tall buildings. Diffraction deforms the signal, as this last is wrapped around the diffracting object. Hence, the waves produced by the diffracting object are present throughout space and behind the obstacle based on the Huygen's principle [38]: *every point on a wavefront acts as a point source for the production of the secondary wavelets. The combination of all wavelets produce a new wavefront in the direction of propagation. The field strength in the shadowed region is the vector sum of the electric field components of all the secondary wavelets that are received by the receiver.*

As described in [39], the effects of reflection, diffraction and scattering are illustrated in Figure 2.2. An understanding of indoor propagation effects is very important when designing a suitable communication system. In mobile communication system, electro-magnetic wave propagation is mostly affected by irregularities arising during its transmission. Irregularities are attenuation, free space loss, shadowing and multipath [40] (cf. Figure 2.3). Such phenomena come from a particular or combined physical abnormality (reflection, scattering, and diffraction) that happens during the wireless signal propagation. Hence, they should be taken into accounting when designing a wireless application.

Recall that attenuation causes over any transmission medium a loss of the strength of the signal over distance. Instead, time varying errors can happen on the received signal due to rapid

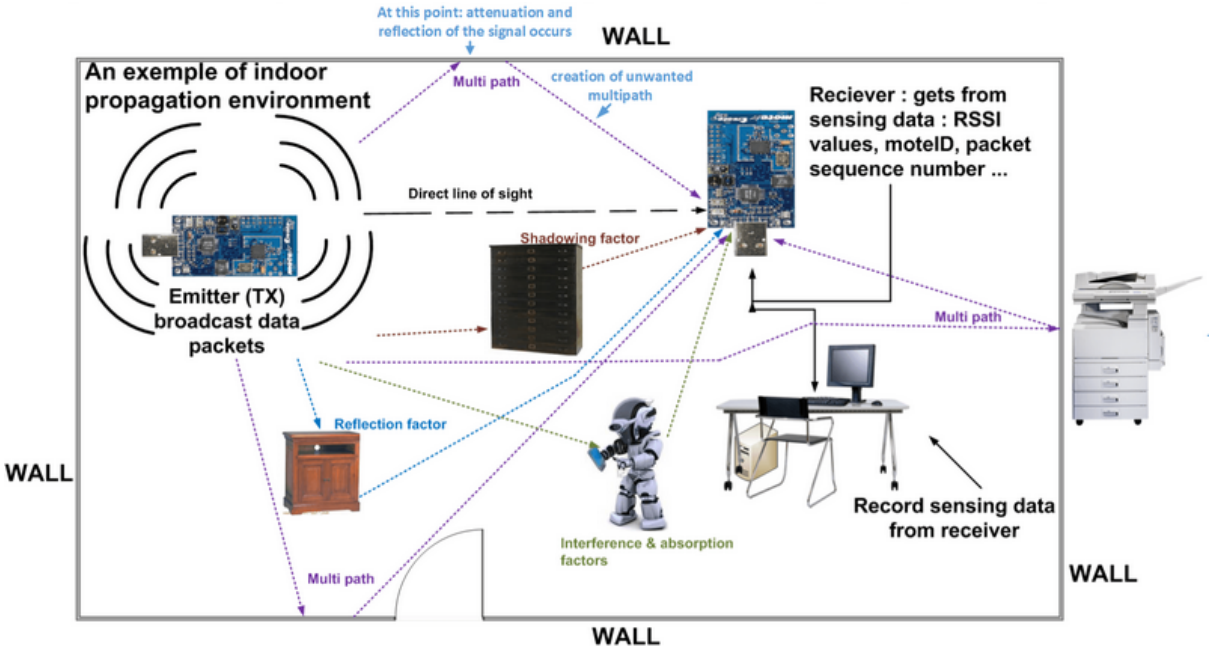


Figure 2.1: An example of our System Architecture with multipath context: a signal received with two or several paths

Figure 2.2: Demonstration of reflection, diffraction and scattering propagation effects

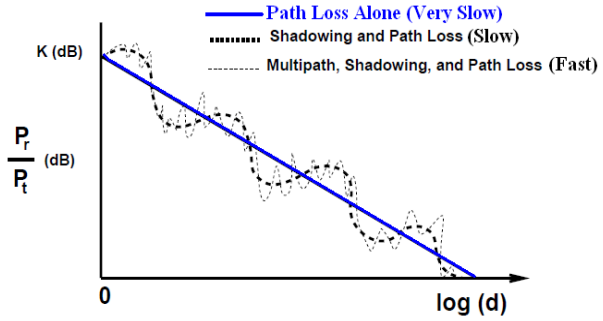


Figure 2.3: Indoor propagation characteristics: Path loss, Shadowing and Multipath in function of the distance [37]

changes on the transmission medium. In an indoor environment where the localization node may be mobile, the location of various obstacles may change over its displacement and therefore imply dealing with complex propagation effects. Thus, the received indoor signal often results on a combination of linear signal loss over distance and changes that happen on the signal. Signal loss over distance is explained by the fact that radio signal is absorbed by building structure like walls, floors, construction materials (e.g. concrete wall) along the receiver path: this is also known as the radio signal attenuation. Instead, changes arising on the signal are explained by sudden and unexpected distortion of the signal. The layout of the room like wall partitions and the presence of obstructing objects, furniture between two communicating wireless devices can occlude the communication link. The effects of these factors can result in fading, reflection, scattering and diffraction.

We have therefore to distinguish both characteristics during signal transmission: short-term or small-scale variations and long-term also known as large-scale variations of the radio wave [25, 41] (cf. Figure 2.3).
Short-term variations of the radio wave consist of large, rapid and sudden changes on the amplitude of the radio signal over a short period of time or distance on a scale of the wavelength λ. Such effects usually labelled as fading, are modeled by small scale propagation models. Short-term variations are either introduced by reflection, scattering objects or the material type of the objects along the radio signal path. Consequently the channel may vary rapidly with the blind mobile node's location and movement. This is especially the case in indoor environment. Small scale propagation models generally evaluate multipath effects and are frequency and bandwidth dependent [42].
The long-term variations on the other hand correspond to gradual changes of the signal strength along the communication path. The received power decreases as the separation distance increases between the emitter and the receiver. Long-term variations are caused in part by diffracting objects and are modeled by large scale propagation models. These propagation models evaluate the received signal pattern/comportment (averaged) over distances largely superior to the wavelength. Large scale propagation models comprise free space path loss, log-distance path loss, shadowing, direction drifts, etc. In addition to distance, large scale propagation models are dependent on environment characteristics and are consequently frequency independent. Short-term variations or fading can be modeled by *ricean distribution* in case of fading with stronger LOS [41]. Works from Zanca *et al.* [5] assert that shadowing is almost constant over long time periods, while fast fading shows rapid fluctuations. Consequently, packets received in different

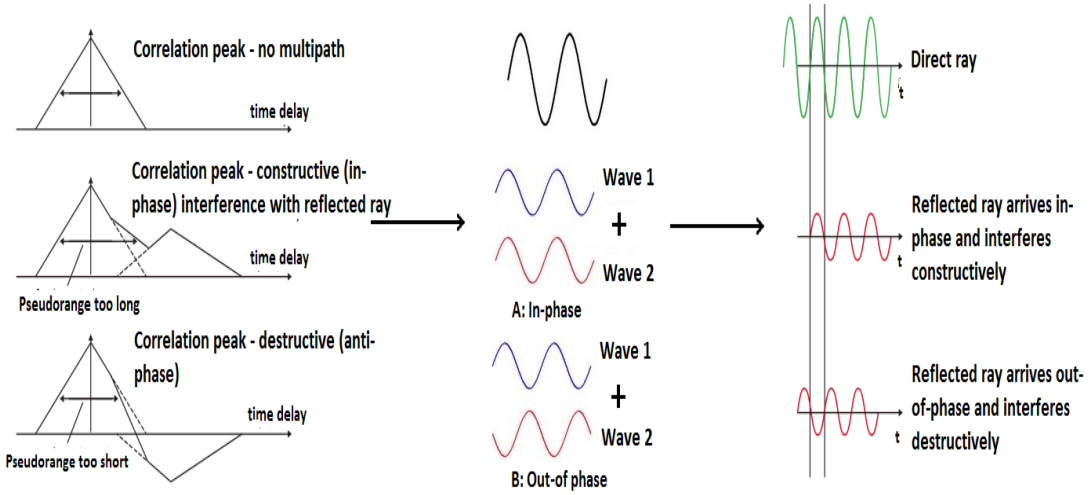


Figure 2.4: Two basics types of multipath: constructive and destructive interference [43]

time sequences likely experiment are affected by shadowing, but almost independent fading. Thus, the fast fading can be averaged out, therefore ignored in slowly changing environments when considering multiple readings from the same static node. In this context, our work is not concerned by fast fading effects.

As illustrated in Figure 2.1, multipath or multiple propagation paths result from the fact that an emitted radio signal arrives at the receiver antenna over more than one path including the direct line-of-sight (DLOS) and the non line-of-sight (NLOS) components. These multiple paths happen during radio signal transmission or radio propagation because of reflections from nearby metallic objects, objets in movement and people moving inside the environment. Consequently, fading and dispersion constitute the principal characteristics of multipath. Recall that individual or mixture of reflection and scattering phenomenon can also convey to multipath as presented in Figure 2.1. As a result of physical interaction from several structures as shown in Figure 2.1, multipath can severely attenuate the received signal, causing the composite signal to be either larger or smaller than the direct signal.
Multipath effects contribute to constructive or destructive interference phenomenon. Constructive interference increases the resulting power and arises when the reflected signal arrives in-phase with the original (line of sight) one. Destructive interference instead decreases the power and takes place when the reflected and original signals are out of phase as shown in Figure 2.4. Let us consider two traveling waves. From Figure 2.4 described in [43] we can notice that if both waves traveling in the same direction with the same frequency arrive in phase, their mutual reinforcement gives rise to a single bigger amplitude. This bigger wave amplitude is a sum of the resulting two waves amplitudes with equal frequencies and phases and leads to the situation of constructive interference. However if the crests of one wave overlap with the hollows of another they produce cancellation and drive to destructive interference. In fact, when the first wave is up, the second wave is down or when the first wave is down and the second is up; the sum of both waves can be less than either the smallest between both (i.e. waves), or can even be zero.

Note that multiple propagation paths can experience both slow and fast fading effects in a narrowband transmission system like Zigbee.

Furthermore, in [37], the authors present a comprehensive description of variation and dissipation in received radio signal power over distance. Variations due to path loss (i.e. large-scale variations) occur over relatively large distances like hundreds of kilometers. Large-scale va-

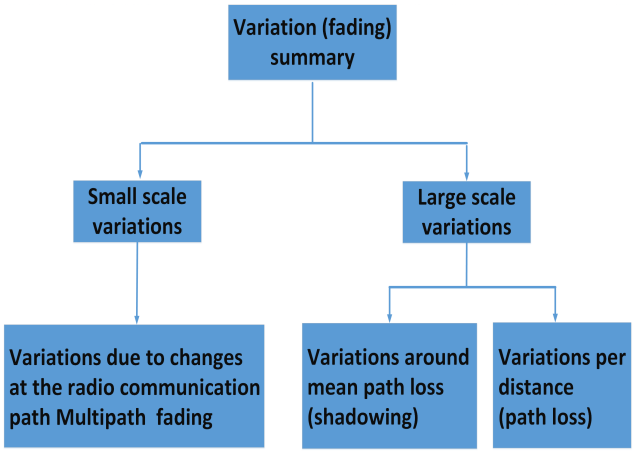


Figure 2.5: Two basics types of variations in indoor environments [25, 44, 45]

riations imply a gradual variation of the received power. Signal attenuation is governed by the entire radio communication path. Shadowing referred to as medium-scale variation happens when a receiving antenna moves toward larger distances ranging from fewer tens to hundreds of meters. However variations due to multipath occur over very short distances in the order of sub-wavelength or wavelength. This is called small-scale variations as presented in Figure 2.5. Small-scale variations count for the rapid fluctuations of the received power, consequence driven by the constructive and destructive interference of the multipath waves. Therefore the indoor radio channel is especially difficult to model. Hence a thorough understanding of indoor radio propagation is required to design accurate and robust localization algorithms.

Due to its random unpredictable nature, modeling the wireless radio channel of the environment is a key tool (black box) that governs the received radio signal in indoor environment. As indoor environment characteristics may differ from each other, a generic and adapted propagation model is needed to examine the detail of how a signal is propagated from an emitter to a receiver. Such propagation model should capture any indoor propagation irregularities. This generic model can be then applied to any specific situation as a standard to follow for reliable indoor communication. We follow earlier works on analytical and empirical validation of propagation models [22], [37]. Each propagation model stands for a mathematical expression, algorithms or diagrams in order to represent the radio characteristics of a given environment. Used to determine the electromagnetic-wave propagation characteristics analytically, theoretical propagation models are based on Maxwell's equations [37]. Maxwell's equations are simplified by path loss models which vary in complexity and accuracy. Several theoretical and empirical propagation models have been developed for indoor propagation [24, 37].

Empirical models identify the radio wave propagation using statistical properties of recognized features based on empirical mathematical formulation from empirical measurements. This makes them more advantageous since they try to separately recognize all factors that influence the environment through the use of few parameters. With the use of parameters, empirical or statistical models capture phenomena missing from theoretical formulas used by theoretical models. The latter instead deals with the fundamental principles of radio wave propagation phenomena. Theoretical or deterministic models apply the principle of physics to the environment features (mostly represented on a database). This sometimes may not be practical or need a frequently update of the database, as the knowledge of the environment is required (building layout/materials and environment geometry). However one should not forget that the propagation of indoor environment often changes. Thus, empirical models contribute more on the design of

generic propagation models and reduce computational complexity. Note that empirical models are widely used to study cellular systems performance via simulation. The following section describes the most used models, mainly empirical models, for estimating the LOS signal path. We present the most used and relevant radio propagation models in order to validate the one to use for our study. We also present results of most contributing works that used such models. The results of current state of the art shed light on existing radio propagation models and inspired our choice. We first describe generalized propagation models, that is large-scale propagation models: the free-space path loss model, the two-ray model (deterministic), the log-distance path loss model. We then present the most used indoor propagation models: log-normal shadowing and the partition and floor attenuation factors models. We end-up with small-scale or fading models. Notice that the log-distance path loss and the log-normal shadowing are empirically proven to work in outdoor and outdoor environment [24]. Typically, the most commonly used propagation model in free space is the Free Space Path Loss also known as Friis and presented in equation (2.2).

$$P_r(d) = \frac{P_e G_e G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{2.2}$$

$P_r(d)$ is the received power and is function of the distance, P_e is the the emitted power. G_r represents the emitter antenna gain, G_r is the receiver antenna gain. λ is the wavelength in meters, d is the distance in meters separating the emitter to the receiver. L is the system loss factor and is not related to propagation. It is usually due to transmission line attenuation, filter losses and antenna losses in the communication system. Generally, $L = 1$ signifies that there is no loss in the system hardware. It is common to select $G_e = G_r = 1$ e.g. for isotropic antennas. An isotropic antenna broadcasts power equally in all directions. The equation 2.2 is only valid in Fraunhofer region. The Fraunhofer region called far-filed region is the region in which the propagating wave looks like a plane wave. Thus the power decays inversely with the distance $d_0 = \frac{2D^2}{\lambda}$ where D is the antenna diameter (the antenna largest dimension) and λ is the radio wavelength. Therefore, the separation distance between an emitter and a radio receiver, d should be much larger in the order of the wavelength than the nearest minimal reference distance d_0 chosen in the far-field zone. Free Space Path Loss predicts the received power to decrease as the square of the emitter-receiver separation distance. In this case of idealized propagation, the received signal power is easily predictable. However, free space path loss model is mostly relevant in outdoor environment. Free space path loss model can address the indoor environment to predict received signal strength only when the emitter and the receiver have a clear, unobstructed line-of-sight (LOS) path between them. In most case, the reference path loss for a reference distance required by other models is computed through free space propagation model.

2.1.1 Deterministic two-ray approximation : ground reflection model

[24, 46]

In the free space model the emitted signal reaches the receiver antenna with only one single path. This model assumes a clear line-of-sight path between an emitter and a receiver antenna. However, as mentioned and described in [39], both Figures 2.1, 2.2 demonstrated that the propagation of radio signal in realistic unsteady indoor environment causes the emitted signal to be interfered between direct and reflected signals. The signal attains the receiver via multiple paths. The two-ray approximation explains such effects by considering the signal to reach the

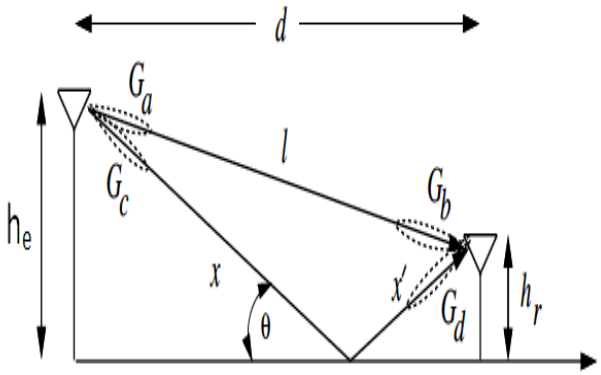


Figure 2.6: Illustration of the Two ray model [37]

receiver in two paths: a direct path and a reflected path. The most common and accurate approximations for signal propagation use the ray tracing techniques [37]. Such techniques consider the wavefront as single particles and determine the reflection effects of the wave by ignoring the scattering which is predicted by Maxwell's equations described in [37]. The two-ray model is the simplest ray-tracing model and a way to model the line-of-sight radio channel. The two-ray model predicts signal variation resulting from a single ground reflection interference with the LOS path. Such interference mostly dominates the multipath effects in outdoor context. The two-ray model is a numerical model that takes into consideration the behavior of high-frequency radio waves in a ray-like fashion. In fact, a more realistic treatment of the path loss takes into account that radio propagations commonly suffer from at least one notable source of interference, named ground reflection [47].

As mentioned, the use of the two-ray propagation model involves a scenario with one reflecting surface [48]. Typically, as demonstrated in Figure 2.6, the reflected path bounces on the ground. In order to receive the reflected signal, a major requirement is that the receiver antenna should be positioned above the ground. The ray-tracing models for path loss require detailed site specific information since they rely upon the geometry and the dielectric properties of the environment in which the signal propagates. Ray-tracing techniques consider the radio signal propagates as a ray between a radio emitter and a radio receiver [49].

They can generally be applied to indoor/outdoor environments. With the ray tracing models the propagation wave related parameters are estimated according to Maxwell's equations. However, such models are more precise under a small number of multipath components and when detailed information on the physical environment characteristics are known. Therefore, such models are environment dependent.

Figure 2.6 shows an illustration of the two-ray ground reflection. A direct ray is represented by the l segment while both $x + x\prime$ segments indicate the indirect (reflected) ray. In fact Figure 2.6 describes a wireless system in which an emitter at a height h_e is sending a radio signal to a receiver device at a height h_r lower than h_e and at distance d. The ground reflection coefficient Γ is defined by the ray angle of incidence θ also called grazing incidence [24]. The value for Γ depends additionally on ϵ_r, the dielectric constant of the ground.

$$\Gamma = \frac{\sin(\theta) - \alpha}{\sin(\theta) + \alpha} \tag{2.3}$$

Where

$$\alpha = \begin{cases} \frac{1}{\epsilon_r}\sqrt{\epsilon_r - \cos^2\theta} & \text{vertical polarization} \\ \sqrt{\epsilon_r - \cos^2\theta} & \text{horizontal polarization} \end{cases}$$

Note that perfect ground reflection imply $\Gamma = -1$. For asymptotical large d, $x + x\prime \approx l \approx d$, $\theta \approx 0$, $G_e \approx G_r$ and $\Gamma \approx -1$ [37].

When the transmitted signal is narrowband, the received power of the two-ray model (at a distance d) approximately in dBm, is a summation from each contributing ray (l, $x + x\prime$) and is given by:

$$P_r(dBm) = P_e(dBm) + 10log_{10}(G_e) + 20log_{10}(H_eH_r) - 40log_{10}(d) \tag{2.4}$$

This equation 2.4 has been demonstrated to closely consent with empirical data. P_e is the emitted power and P_r, the received power. G_e represents the radiation pattern in the LOS direction and is defined as the multiplication of emitter, receiver antenna field radiation: $G_e = G_aG_b$. Instead, G_r is the radiation pattern of the indirect ray of length x and $x\prime$. $G_r = G_cG_d$.

Compared to the Free-Space model (cf. equation 2.2), the two-ray ground reflection model (cf. equation 2.4) presents a faster power loss as the distance increases: the received power attenuates to the fourth power of the distance. The two-ray model is known to better predict the received power over long distance compared to free space model [24, 47, 50]. Because of oscillation due to the constructive and destructive combination of the two-ray, free space model is mostly used when the distance d is smaller compared to the two-ray approach. Distance is of the order of several kilometers for mobile radio systems with tall towers heights exceeding 50 m in urban environment. However in most mobile communication systems, the maximum distance between any radio emitter-receiver is only a few tens of kilometers and the earth is assumed to be flat [24]. Two-Ray Ground models are applied in most common network simulators (e.g. NS-2.34, NS-3.11, OMNeT++) [47]. However, experimental measurement results obtained in outdoor, on the road and presented in [47] have demonstrated that Free-space and two-Ray Ground models are not able to capture complex path loss effects at small and medium transmission distances. At small distances ($\leq 140 - 200$ m) the prediction errors alternate between underestimating and grossly overestimating RSS values i.e. between -5 dB and $+10$ dB. While, at mid distances, i.e. 600 m, distance predictions by both simpler models consistently underestimate RSS values by more than -5 dB. Works in [37] additionally state that the two-ray model is not typically a good model for indoor environments as it characterizes signal propagation in isolated areas with few reflectors, such as rural roads, along highways as well as over water.

Free-space and two-Ray Ground models are thus only relevant in scenarios where reflections on buildings do not dominate radio propagation effects [47], like in suburban environments.

2.1.2 Multipath fading characterization : the Rayleigh and Ricean distribution

As mentioned, an emitted signal running into reflection and scattering can result to several paths at the receiving antenna side. Each electromagnetic wave path can have its own delay. An electromagnetic wave phase and amplitude may vary depending on the receiver node movement as well as on the collision impact from local object in the environment. Such phenomena can conduct to several multipath fading types as shown in Figure 2.7.

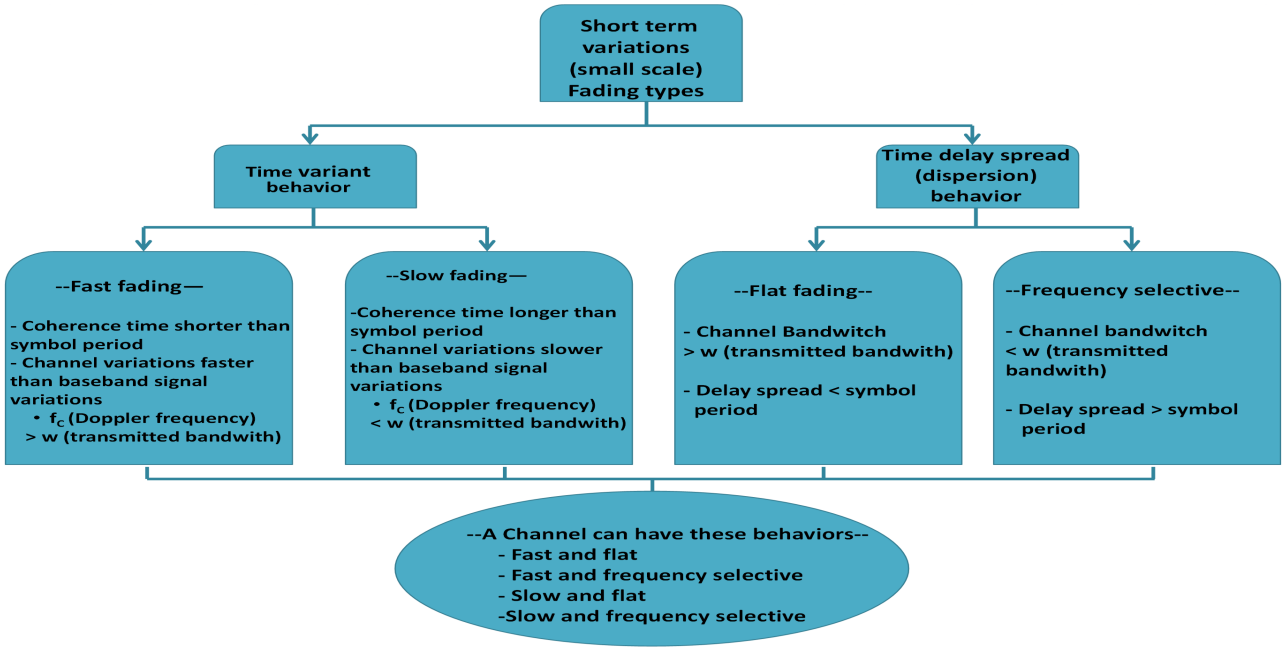


Figure 2.7: Illustration of multipath fading types [25, 44, 45]

Described in [44, 45], two types of multipath fading are presented. Time variant introduced by motion and changes in propagation paths; and time spreading behavior due to signal dispersion, for instance named multiple paths. Let's denote by f_c the Doppler frequency and it measures the time variance of the channel characteristics. Doppler frequency is based on the carrier frequency and the relative speed of any movement along emitter and receiver communication path. Doppler effect can be explained by the fact that a single signal (or sinusoid) is received on a receiver as a summation of three signals (or sinusoids). Each of the three signals is the sum of the carrier frequency (named f_0) and the Doppler frequency of each sinusoid (i.e. f_c). The Doppler bandwidth (i.e. range frequencies without high attenuation) is inversely proportional to the coherence time (i.e. channel change rate) and both specify for short term variation the time variation of the channel [25]. The coherence time is the channel change rate in time domain over which the channel is not invariant. Fast and slow fading focus on change of the signal during wireless communication when considering the changing rate of the channel. Such effects are in relation with the velocity of the mobile node, the baseband signaling (the channel and the transmitted signal) [25, 44, 45]. Fast fading is due to fast motion speed and assumes that weakening on the strength of the signal do not vary as a function of frequency. With fast fading, the channel may change drastically in amplitude and in phase several times during the usage time [25, 44, 51]. Instead, slow fading which happens in most communication system assumes the channel varies slowly in amplitude and in phase or even constant over time during signal transmission. In this case, we deal with the coherence bandwidth also called the bandwidth of the channel and the signal bandwidth (i.e. w).

Multipath fading types referred to as Rayleigh fading types or Rayleigh distributions are used to model scattered signals that reach a receiver by multiple paths. The Rayleigh distribution assumes a received signal is composed of a sum of severals large waves with random individual amplitudes and a uniform distribution of phases. Note that, the central limit theorem assumes the sum of a sufficient number of random variables progress closely to a normal distribution [25]. Thus, when two components of the received signal, example, x and y, are normally distributed, the received signal is a complex Gaussian variable. In other terms, when x and y are independent

Figure 2.8: Illustration of shadow fading (i.e. shadowing) and Rayleigh fading [44]

Gaussian random variables with zero means and variances $\sigma_x^2 = \sigma_y^2 = \sigma^2$, the joint probability density function (pdf) of the Rayleigh distribution can be modeled as [41], [52]:

$$p(r) = \begin{cases} \frac{r}{\sigma^2} e^{\left|-\frac{r^2}{2\sigma^2}\right|} & r \geq 0 \\ 0 & r \leq 0 \end{cases}$$

Where $E(r^2) = 2\sigma^2$ is the total power of the received signal and the envelope amplitude of the received signal which is Rayleigh distributed is given by: $r = (x^2 + y^2)^{\frac{1}{2}}$. This amplitude is independent from the phase that is normally distributed. $r^2 = 2\sigma^2$ is the power of the signal with exponential distribution. Such power is also known as mean-squared value of the distribution.

Rayleigh distributions describe the statistical time varying nature of the received envelope of a flat fading signal [41, 53]. Flat fading (cf. Figure 2.7) is the most common type of fading in which the multipath structure of the channel are preserved at the receiver, however the strength of the received signal changes with time, due to fluctuation in the gain of the channel caused by multipath [53].

A Rayleigh distribution is mainly applied for the use in statistical communication theory. In fact, in most wireless communications, the envelope of the received carrier signal is Rayleigh distributed or Rayleigh fading [52]. Recall as presented in Figure 2.8 the signal displays different fading characteristics based on the density of the scatter in the multipath case. Hence, when the dominant signal becomes weaker, such as in a dense scatter with non LOS case (i.e. NLOS), signal is modeled with Rayleigh distribution while Ricean distribution is used to model fading with a stronger line-of-sight. In fact, when a dominant signal component comes out at the receiver side with many weaker multipath signals, it gives rise to the Ricean distribution. In addition, as a dominant signal becomes weaker, the composite signal approximates a noise signal which has an envelope that is Rayleigh. Thus, when the dominant component fades away, the Ricean or Rice distribution becomes a Rayleigh distribution. Other implementations of the Rayleigh distribution exist. Those are propagation models that give more control over the extent of the fading as the Nakagami distributions and they can be reduced to Rayleigh distributions [41].

To summarize, free space path loss model is too simple and requires clear line-of-sight between an emitter and a receiver antenna. In fact, according to this model, the strength of a radio

signal decreases with farther distance. However, in most indoor cases, the signal strength is coupled with random fluctuations on the amplitude of the signal which depends on obstruction along the signal path. Those random variations imply unpredictable behaviors on the measuring parameters (e.g. the received signal strength) in the concerned area while greatly complicating at the same time the analysis of the indoor wireless link behavior. Deterministic ray tracing models require site-specific information and complex mathematical operations. Rayleigh distributions and further implementations do not fit for the use in indoor environment as mentioned above due to their high-level analysis requirements. In this context, efforts were devoted by researchers for understanding the behavior of wireless links. Propagation models that best consider complex behavior of indoor propagation are then proposed. Seidel and Rappaport [22, 24] introduced the log-distance path loss model and the log-normal shadowing. The former model (i.e. log-distance path loss) easily represents path loss as a linear function of the distance. The latter (i.e. log-normal shadowing also known as log-normal path loss model) considers possible additional random amplitude fluctuation on the received signal based upon measured data. The next paragraph summarizes both models, being commonly used as large scale models to describe signal strength.

2.1.3 Indoor propagation models

The log-distance path loss model According to [24], both empirical and theoretical-based propagation models have demonstrated the average received power to decrease logarithmic with distance either in indoor or in outdoor environment. The path loss represents a positive quantity, the difference in dB between the effective transmitted power and the received power. The distance of reliable communication between a wireless emitter and receiver can be predicted by equation 2.5 using a path loss exponent: the strength of a radio signal decays exponentially with distance.

$$\mathrm{PL}(d) = \bar{\mathrm{PL}}(d_0) + 10\eta \log_{10}(d/d_0), \tag{2.5}$$

Where $\bar{\mathrm{PL}}(d_0)$ is the reference power loss (in dBm) at reference distance d_0 (in meters) for large scale path loss models. In other words, the reference power loss is assumed to originate from free space propagation at a shorter reference distance d_0 from an emitter. d_0 is chosen in the aforesaid far-field antenna region. Typically, the value for the reference distance d_0 is equal to 1 m in indoor and from 100 m to 1 km in outdoor. The value for the reference power is usually computed either through measurements (taking the average received power) at the reference distance d_0 (i.e. from many closed-in points located at a radial distance d_0 from the emitter); or predicted from the free space equation discussed above. η is the path loss exponent. It gives the rate at which the path loss increases with the desired distance d (in meters) that separates the emitter to the receiver antenna. The value of the path loss exponent η is environment dependent and is estimated with empirical data. In free space, $\eta = 2$. Consequently, the log-distance path loss model is based on the free space path loss while accounting for more environmental factors [42]. However, the log-distance path loss model does not consider environment occlusions effects that imply variations on the radio signal path between an emitter and a receiver and thus causing shadowing. Siedel and Rappaport have attempted to overcome this issue by proposing the empirical log-normal shadowing model [22].

The log-normal shadowing During its transmission, the radio wave deals with different obstructions like building type, building wing, tunnels, indoor furniture pieces and the way

such furniture is arranged. Such factors may weak the signal while causing at the same time shadowing. The shadowing effects are taken into account by considering a zero mean and Gaussian random variable for the equation of the log-distance path loss propagation model. Seidel and Rappaport have observed and demonstrated in a former work in [22] that path loss is log-normal distributed when assuming the distribution of gradual large-scale variations (named large scale path loss) as shown in equation 2.5. Even if, their measured data do not always present a log-normal distribution, the mean path loss exponent η and the standard deviation σ (expressed in decibels) are computed according to equation 2.6 expressed as:

$$\mathrm{PL}(d) = \bar{\mathrm{PL}}(d) + \epsilon\,, \tag{2.6}$$

or

$$\mathrm{PL}(d) = \bar{\mathrm{PL}}(d_0) + 10\eta \log_{10}(d/d_0) + \epsilon\,, \tag{2.7}$$

$\mathrm{PL}(d)$ tends in this case to a log-normal distribution, $\bar{\mathrm{PL}}(d_0)$, η, d and d_0 have a same meaning as in equation 2.5. $\epsilon \approx N(0;\sigma)$ is a zero mean log-normal distributed Gaussian random variable with the standard deviation σ and accounts for the random effect of shadowing. ϵ models the path loss variation due to shadowing through all location at distance d separating the receiver to the emitter. The standard deviation σ represents quantitative measurement of the accuracy of used path loss model for an obstructed coverage area. In addition, the values of both the path loss exponent η and of the standard deviation σ are environment dependent and are estimated in decibels (i.e. dB). Hence, this capability imply the log-normal model to more accurately predicts path loss as a function of the distance. Notice that, measured received signal strengths from different locations being at equal distance from an emitter are considered to be independent and identically distributed (i.i.d.) normal random variables. Moreover, measurements data for 900 MHz frequency between a wireless emitter and receiver located on the same floor within four buildings have demonstrated that the log-normal shadowing model matches well empirical data for values of $\eta = 2.76$ and $\sigma = 12.9$ [22]. Instead, for measurements where the emitter and receiver are on different floors, the path loss is more severe, and the standard deviation is reduced. These parameters values can be used in a model for a first-order prediction when no specific building information is available. However, one should not forget that the reported values for η and σ are particularly limited within a coverage environment. Thus, for a more accurate indoor propagation model, building features around emitters and receivers may be considered.

Modification of the log-distance model with addition of attenuation factors: partition losses between floors Additional path loss models have been studied by researchers with consideration of individual losses between a wireless radio emitter and a receiver path. These losses are introduced by obstructions including the number of walls, floors and doors travelled by a radio signal. Obstructions arise also from a turn corner in corridor, aluminum materials. Theses enhancement are the first step for including detailed site information to improve propagation prediction. Each additive path loss model first estimates each individual attenuation loss factor based upon measured data before adding them together to constitute the overall losses of the targeted environment. More representative models that add attenuation factors for log-distance path loss model are presented below. The floor attenuation factor prediction model proposed by Seidel and Rappaport [22] in multi-floor environment evaluates the mean path loss as a function of the distance when considering the mean path loss exponent as a function of the number of floors between an emitter and a receiver. In addition, for a particular building type, a constant floor attenuation factor named FAF is added. FAF is function of

the number of floors and building type. FAF is added to the mean path loss predicted by the log-distance path loss model at the same floor and is given by:

$$\mathrm{PL}(d) = \bar{\mathrm{PL}}(d_0) + 10\eta_{samefloor}\log_{10}(d/d_0) + FAF, \tag{2.8}$$

The value of FAF (dB) varies based upon the number of floors separating the emitter to the receiver. This equation 2.8 should help to empirically predict the attenuation caused by floors as the floor attenuation factor between different floors is shown to relatively differ by at least 3 dB [22]. Such result is obtained when considering two types of building with reinforced concrete. The first building has ten years old compared to the second building. However, it is noticed that the average FAF is not a linear function of the number of floors between an emitter and a receiver. In fact, different floors can cause different amount of path loss. It is also highlighted that factors such as multipath reflections may affect the path loss. Instead of considering the attenuation factor to vary based upon the number of crossing floors within a building; other models consider the same attenuation regarding the traversed walls/floors which is multiplied by the number of walls or floors that separate an emitter to a receiver. Seidel and Rappaport [22] proposed a model which takes into account the number of concrete walls and soft partitions attenuation factors as presented in equation 2.9 shown below:

$$\mathrm{PL}(d) = 10\eta\log_{10}(\frac{4\pi d}{\lambda}) + pPAF(dB) + qWAF(db), \tag{2.9}$$

Where PAF and WAF are respectively the individual soft partition attenuation factor and the individual wall partition loss factor. Both values are respectively multiplied by the number of soft partitions and the number of concrete walls along an emitter and a receiver path. What should be noted with the use of this model is that any kind of concrete support column encountered between an emitter and a receiver communication path will be considered as introducing the same attenuation factor. Such consideration holds even though the components and the structures may be different from one wall to another wall, from one partition to another one. This large scale path loss model is the one used in [54] in order to consider the attenuation terms due to losses from the walls and floors while mapping the measured RSSI to the distance. In [55], Keenan and Motley studied additional propagation effects for stairs being present along an emitter-receiver path in indoor environment and expressed in equation 2.10:

$$\mathrm{PL}(d) = \mathrm{PL}(d_0) + 10\eta\log_{10}(d) + fFAF(dB), \tag{2.10}$$

f is the number of floors/stairs and FAF, the floors attenuation factor. Another new empirical model has later been introduced by Cheung *et. al* [56]. The authors model is based on the equation 2.9 and includes more features such as the distance dependence of path loss exponent, the angle dependence of the WAF, reflection and diffraction. The authors hypothesis is built on the fact that prediction accuracy can be poor in certain parts of a building and especially at larger distances from the emitter. Several other scientific contributions on the modification of the log-distance path loss model including hallways to a main corridors can be found in the literature [24]. In addition, it was observed that due to differences in the type of construction materials and the shape, the path loss within a home can be smaller than the one encountered in an office building with multiple floors. After introduced the existing propagation models and compared them, in the following, we present the state-of-the art of proposed RF-based localization algorithms.

2.2 Localization algorithms

2.2.1 Indoor localization technologies

Several localization technologies have been developed so far. First of all, there are cellular networks which directly incorporate the concept of localization and offers the advantage of working with a common standard for the description of the position [57]. positioning technique in cellular-based networks operates in a licensed band and combines the position of cellular phones to a belonging cell [58]. When using cell-ID positioning techniques, the identification of the cell using the Cell-ID method is realized by combining the position of a cell-phone carried by a person to the coordinates of the antenna that relays the communication. Available measurements are the propagation time of the signal or the angle of arrival of the signals from the mobile station at several base station Node.

A localization system based on cellular signals, such as Global System for Mobile communication (GSM) works in this way. Through time or angle measurements from three closest relay and using trilateration or triangulation lead to define an area in which the mobile cellular phone is located (i.e. a GSM emitter carrying by a person). However localization is less precise when the coverage area of the cellular network becomes larger. For example in rural environment the localization accuracy with Global System for Mobile meaning (i.e GSM) can reach 5 kilometers while in urban areas it varies from 50 to 300 meters. The availability of small cells for indoor coverage has enable cellular-based indoor localization. As a cellular-based indoor localization, GSM-based indoor localization system mostly use signal-strength fingerprints method. Otsason *et al.* [59] present a GSM-based indoor localization using fingerprints and K-nearest neighbors to localize a mobile station to the closest belonging cell. An accuracy ranging from 2.48 m to 5.44 m in multi-floor buildings has been obtained. In [60] another GSM-based indoor localization is presented with the use of wide signal-strength fingerprints and 6-strongest cells. From the authors, this GSM-based fingerprint indoor localization is able to identify a floor correctly in up to 60% cases and within two floors in up to 98% cases.

The Universal Mobile Telecommunications System (UMTS) standard named 3GPP is another technology based on cellular network. Works from *et al.* [57] compare the performance of UMTS and WLAN using radio frequency-fingerprinting in location estimation in an indoor office environment of size 20 m×15 m. The location estimation architecture uses four 3G small cells and four WLAN-APs. The radio map is created through fingerprints collected for a specific period of time at predefined points. Each point is taken within a grid of one meter using access point signal visibility and RSS as metrics. Position estimation of the mobile smartphone client during the online phase is performed by a weighted k nearest neighbors algorithm (WKNN) that uses deterministic Euclidean Distance norm. Precisely measured RSSI or either access point signal visibility is compared to the previous calibrated radio-map. Then calibration points whose metrics have the smallest distance to the current measurement, are used to estimate the position. Experimental results report an accuracy with UMTS comparable to the one achieved with similar WLAN-Testbed. Both methods reach an accuracy between 5 m and 7 m at 80% confidence level. Achieved results are almost independent of the used metric (RSSI or visibility), of the distance norm or the amount of sample collected in the online phase. Thus UMTS small cells are good alternative or complementary to WLAN for indoor RF-Fingerprinting with the use of smartphone. In fact, the same hardware can be used for traffic and for localization purposes. Additionally UMTS does not use frequency hopping or power control, as it is the case in most WLAN installations for interference reduction. However, the cell coverage of both GSM and UMTS are still larger to be applied to room level. GSM or UMTS based indoor loca-

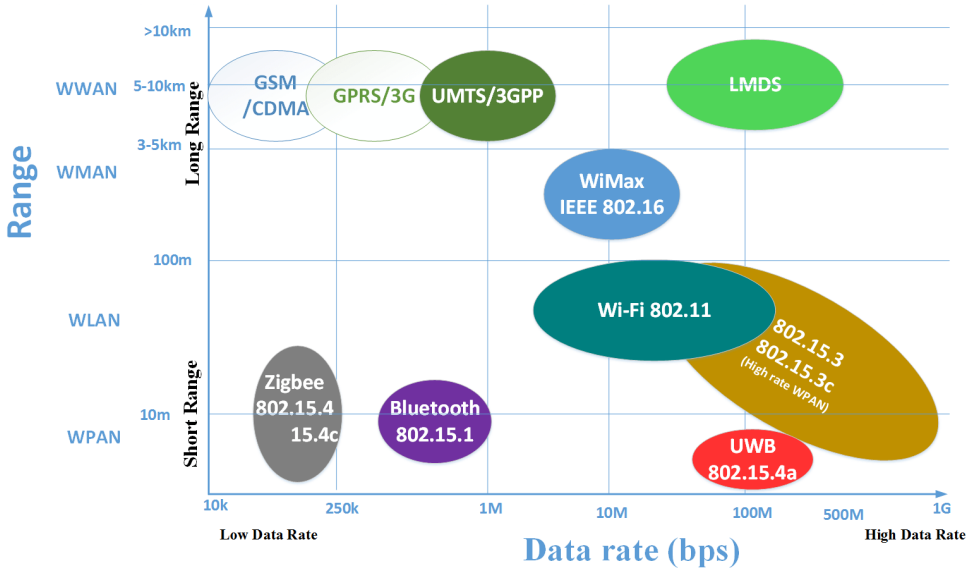


Figure 2.9: Overview of the range of IEEE 802 wireless technology [62, 63, 64]

lization algorithms can not leverage the lower rate and low power feature of localization based on Low-Rate Wireless Personal Area Networks. Fine grained location accuracy is more difficult to achieve with GSM or UMTS than localization based on IEEE 802.15.4 wireless networks. Figure 2.9 presents a classification of wireless networks according to the range and the data rate while Figure 2.10 shows a summary of existing location system described in [61].

Furthermore, other research efforts have exploit the radio system for localization purpose. Radio Frequency Identification (i.e. RFID) is one of the localization system already used for tracking or access control [65], [66]. RFID is an identification system that uses radio waves to communicate between a reader and an electronic tag (i.e. RFID tag). A tag can store information like its identification number, manufacturer information, product type and price [67]. At any time, a reader can retrieve and modify information stored on a tag. A tag is incorporated or attached to an object. There are two types of RFID tag: the passive and active one. This technique of radio identification allows contact less communication of data contained in a tag via radio links. In most case, communications between tags and readers do not involve greater energy consumption on a tag as power is supplied by induction from the reader. Examples of application can be found in various areas including in logistics for delivery of merchandises/goods tracking (e.g. by Wall-Mart), retail stock management or for the usage by some automatic tools booths on highways. However, this RFID type of localization infrastructure is not suitable in our case. In fact, it is not possible to store complicated contents or dynamic information such as RSSI measures which is instead available with wireless sensor networks.

Another approach of localization is based on Wireless Fidelity (Wi-Fi). Wireless Fidelity is a data transmission technology via radio waves and standardized on 802.11a/b/g. Wi-Fi is mostly employed by Wireless Local Area Networks (LANs) which have traditionally been used to provide only data services via radio waves involving an emitter and a receiver at speeds up to 54 Mbits. Most proposed Wi-Fi-based localization algorithms use fingerprinting or location pattern matching. As discussed in Chapter 1, a fingerprint or a pattern matching based localization uses the assumption that a radio frequency signature of any environment is correlated to a location of a target in that environment. A target also called blind mobile node can be any type of object or person that needs to know its location. In most case, these mentioned approaches only

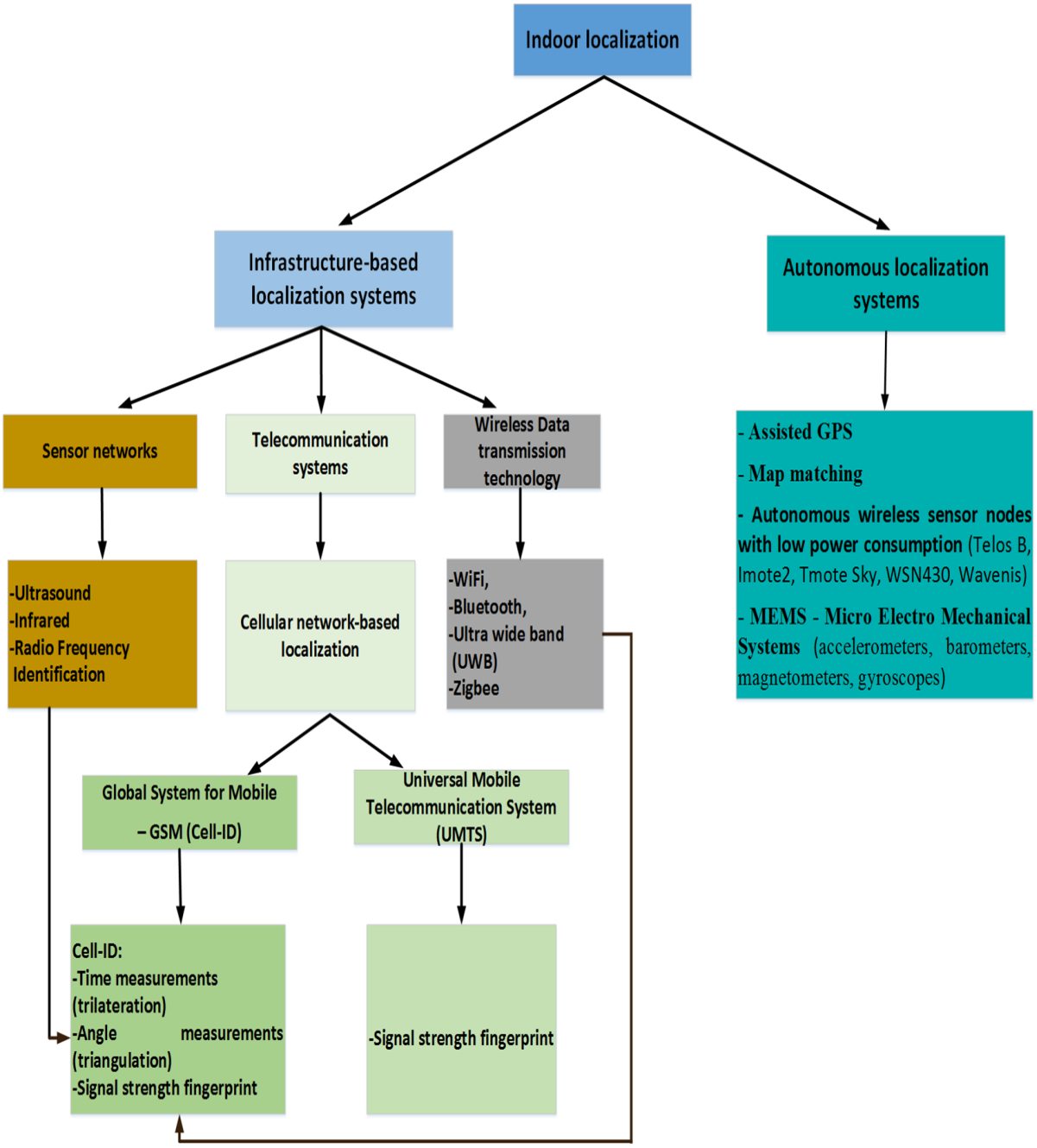


Figure 2.10: Overview of localization systems

require a site survey of the environment features. For instance, the measurements of the signal strengths of several locations to store in a database. During the online phase, a comparison using pattern matching is made between the training set and the current measurements parameters to determine the blind node unknown position. For example, years ago Google was using this fingerprint technique by spotting Wi-Fi networks to feed into their localization database before being stopped for privacy reasons. A fingerprint technique can be an accurate and precise indoor localization algorithm. However, it requires performing a fine-grained and exhaustive offline training phase, each time there is a change in the concerned environment; this is time consuming. In addition, Wi-Fi technology uses a high data rate that is not compatible with low power system/networks including sensor networks. In addition, this type of localization technique is not always accurate than those based on radio propagation model. Methods-based on propagation models use mathematical approaches to estimate the distance between wireless nodes [68].

In [19], in a floor building with dimensions of 43.5 m by 22.5 m, localization results report an accuracy of 4.3 m for the one using a propagation model against 2.94 m for the empirical fingerprint method when considering 50th percentile of the signal strength sample. Instead the 25th percentile of the signal strength sample provides a localization accuracy of 1.86 m for the propagation method compared to 1.92 m achieved by the fingerprint method. Thus, fingerprint-based localization technique depends on the database update frequency which changes in time according to environment characteristics. Several works are interested by the update of the database. Kumar *et al.* [69] have implemented a *Bayesian* method for positioning inside a wireless local area network. The authors demonstrated the localization accuracy to degrade with the signature age. The authors have additionally shown by integrating a *Kalman* filter to suppress the noise, the improvement of positioning accuracy by 90% instead of 80% estimate corresponding to 2.5 meters without the use of *Kalman* filter. Lin *et al.* [70] have paid attention on the issue of the instability of fingerprinting in indoor environment of size 12 by 50 meters space. The authors study the impact of the antenna orientation, foreground obstacles and nearby moving obstacles. Experiments have shown that the RSSI variance is high when the wireless node is held in hand while this variance is smaller and the RSSI is more stable when the wireless node is placed on a chair. In addition, the localization error varies from 1 to 3 meters depending on the orientation of the receiving wireless node antenna. It came out in order to eliminate the need to often repeat calibration, an antenna with a uniform radiation pattern should be used. The receiver node antenna also needs to be hold at the same orientation in order to keep a linear localization error. The authors also highlight that the landmark density influences the localization accuracy but have no impact on the signature instability.

Instead, Ultra Wide Band (i.e. UWB) [71] is a promising low-power, low-cost and high accurate technology. It is a recent trend in propagation time measurement and it follows the IEEE 802.15.4a standard. Ultra Wide Band technology works by sending precise timed digital pulses of short duration on a carrier signal across a very wide spectrum. Thus, UWB emitter and receiver must be coordinated to send and receive pulses with a good accuracy. Through UWB, a signal pulse is transmitted at a very high speed from 1 to 27 Mbps and the radio energy is spread over a very wide frequency band ($\geq$ 500 MhZ). The low power spectral density limits the interference potential with conventional radio systems while the high bandwidth can allow very high data throughput for communications devices. These features of UWB allow good location accuracy of the order of decimeter [72], [61], [73]. In addition, the shortness of the measurement pulse which can occupy several GHz of spectrum, but only nanosecond scale, makes the UWB a very advantageous radio technology that is resistant to multipath compared to the aforementioned technologies. However notice that the high signal propagation speed and the required high

sampling frequencies can imply a limited range. Note again that UWB is an attractive technology for ad-hoc networks and is mostly used in body area networks [74]. However, as we rely to RF RSSI-based indoor localization, UWB technology is not suitable. In fact, UWB signal needs time synchronization among nodes [75]. Note that, RSSI metric is exploited in UWB in most case in hybrid context with time measurements techniques. According to [76], the unique very large bandwidth characteristic of a UWB signal is not exploited to increase the accuracy in the case of RSSI-based localization algorithm. This latter is very sensitive to the estimation of the characteristics of the channel. In addition, accurate angle estimation using the large bandwidth of a UWB signal is very challenging due to scattering from objects in the environment [76]. Thus, with UWB, only the time-based approaches provide precise location estimates.

Another short-distance wireless technology is Bluetooth, a standard for connecting handheld wireless devices with other similar devices. It is also one of the most promising and cost effective candidate to provide short range (i.e. 10 m) wireless connectivity for mobile users. The building block of the bluetooth network is a piconet and a small cluster of devices that share a common channel. Devices must be synchronized to the same clock and must adopt the same frequency hopping sequence. A piconet is a small network that is instantly and automatically created between multiple Bluetooth devices within the same radius and using a star topology of type a "master" and several "slaves". The main features of the Bluetooth technology are robustness, low complexity, low power consumption. Using bluetooth to locate any target can be expensive [77], since Bluetooth operates in short range. The area must be then equipped with a large number of bluetooth devices. A example of Bluetooth-based indoor localization system is proposed by Fisher *et al.* [78]. The proposed localization system can localize any bluetooth devices with an accuracy of 1 meter and needs the use of additional hardware on the base station. However, Bluetooth technology requires clock synchronization among Bluetooth devices. This prevent its use in our localization system.

Last but not least, there is ZigBee a short-range wireless data transmission standard that uses low power, low throughput. Zigbee has enabled low energy consumption and low cost products compared to Bluetooth, UWB and Wi-Fi technologies. Adopted in 2004 by the ZigBee Alliance [79, 62], Zigbee is originally designed to interconnect embedded devices/units like sensor nodes, control units with energy constraints at low data rate of 250 kbit/sec. Nevertheless, even if ZigBee can be a good choice in many low-power communication applications like building automation (e.g. automatic thermostat and light control), it is less popular in the areas of wireless sensor networks. In fact, the Zigbee radio module is known to have the highest power consumption due to packet communications. In addition, Zigbee is built on the basis that the surround devices may have a constant source of energy. Zigbee does not also provide any power-saving mechanism for routing nodes and consequently the lifetime of battery-powered ZigBee network (i.e. network of devices) is limited to a few days [80]. Therefore the current Zigbee specifications are not approved in the field of sensor networks. Instead, the usage of sensor nodes have become key elements in any system that requires information from outside the surrounded environment.

To improve the energy consumption new MAC protocols appeared alike power-saving X-MAC protocol, B-MAC and duty-cycling Contikimac [81] and these are more attractive for the use in sensor networks. With such proposed MAC protocols, there is no random back-off then emission of the entire frame but successive sending of preambles when an emitter wakes up until the receiver responds with an acknowledgment and ready to receive the complete frame. Other sensor nodes can go back to sleep once they saw that the package was not intended for them. The send of successive preambles containing the address of the receiver to wake up is not compatible with the current IEEE 802.15.4 standard. However Contiki operators [80] state

that this mechanism is easy to incorporate into future versions of the IEEE 802.15.4 standard. Currently the CC2420 [82] radio layer of sensor nodes and later versions are able to send packets conform to the specifications of the physical layer of the IEEE 802.15.4 standard. These features of wireless sensors networks make them more reliable to be used with radio frequency technology. Indeed, sensor networks are communicating objects that can provide services themselves. Such services are provided individually or in cooperative way to form a network while offering at the same time simultaneous services. For example coupled with a localization algorithm, a sensor network determines the precise positions of network components that are then transmitted to the home application control. Consequently, the building may be well able to detect the presence of a user and adapt its behavior depending on the position of the user and the environmental conditions. The next section presents the families of RSSI-based localization algorithms.

2.2.2 Radio Frequency (RF) families of localization approaches

Localization techniques mainly used two families of approaches: range-based and range-free algorithms. Range-free algorithms use information coming from network properties such as the number of hops between wireless nodes in a network, the radio proximity between two wireless nodes, the radio connectivity, the average number of neighboring nodes. Precisely, range-free RF-based localization algorithms rely on the simple information that if two nodes can communicate together, they are likely close or less than a certain distance defined by the communication range. Combining such proximity data can lead to an acceptable accuracy, especially in dense environments [28, 31, 83, 84]. In this thesis, we consider the range-based algorithms. However, we take a look on range-free localization algorithms using few examples in order to argument the reason that such types of algorithms are not suitable in our case.

2.2.3 Range-free algorithms

In range-free algorithms, node location is typically determined using either the network connectivity or the radio signal strength information coming from immediate neighbors. Range-free algorithms can be categorized into two types: those exploiting the network connectivity (cf. Figure 2.11) and those using the signal strength measurements to construct a protocol through which a location is estimated [85, 86, 87].

The first category of range-free algorithms exploits the network connectivity using for instance, the number of hops rather than absolute euclidean distances. The number of hops indicates how close a node is with respect to its neighbors. The blind node location is then estimated with the use of localization algorithms like trilateration. The most famous proposed localization algorithm in this category is Ad-hoc Positioning System (i.e. APS) introduced by Niculescu *et al.* in [29]. APS is a range-free and distributed algorithm. It builds a coordinate system based on the distance in number of hops that separates two neighbors. The algorithm is implemented in Ad-Hoc networks in which a node can only communicate with its immediate neighbors. So, nodes which are not direct neighbors of a landmark (i.e. node with known position) infer their position with a respect to a landmark using a forwarding distance propagation method. Landmarks coordinates are obtained using the global positioning system. An immediate neighbor of a landmark can estimate its distance to the landmark using signal strength measurements. Then, any second hop neighbors of landmarks are able to infer their distances (to the landmark) employing either the distance vector-hop propagation (DV-hop) or the distance vector-distance (DV-distance) or even the Euclidean propagation method. The rest of the network pursues in

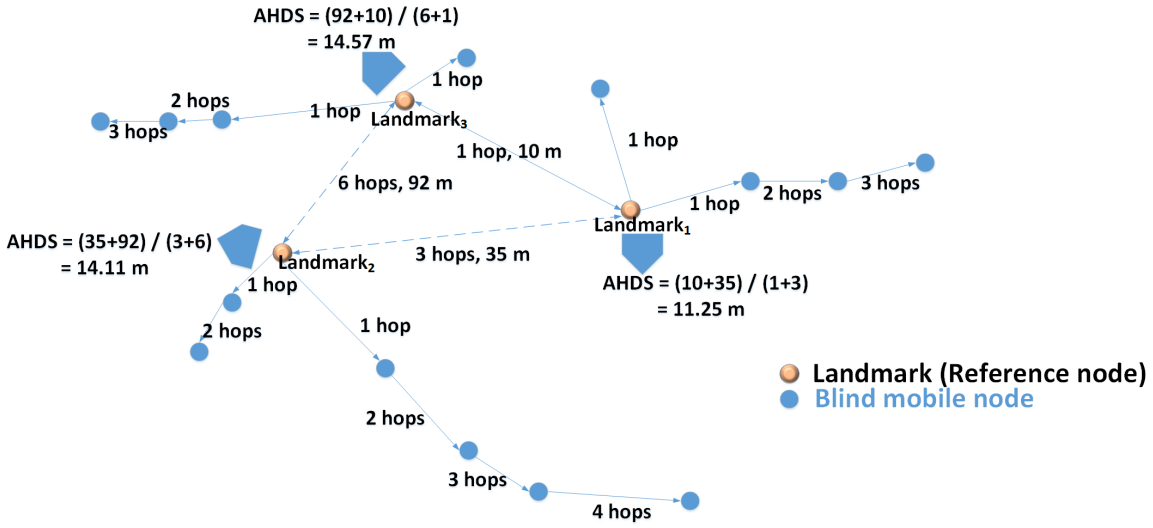


Figure 2.11: The DV-hop algorithm principle

a controlled flood manner originated at the landmark [29]. DV-hop method [88] propagates the distance in number of hop. DV-distance propagates the current measured distances estimated from the received signal strength. Euclidean propagates the geometrically computed Euclidean distances between nodes.

The DV-hop method works like distance vector exchange and shares distance information in number of hops between nodes in an ad-hoc network. Each wireless ad-hoc node named landmark maintains a table: X_i, Y_i, hop_i and then exchanges it with neighboring nodes. (X_i, Y_i) represents the coordinates of a landmark i. Hop_i is the length of the corresponding path or the number of hop to that landmark i. Hop_i is initially set to 0.

During the network setting up, each receiving wireless ad-hoc node in the network adds to that path length the measured hop as presented in Figure 2.11. The modified message is then broadcasted by any wireless ad-hoc node only if the current path length is less than the previous one. In this way, regular wireless ad-hoc nodes as well as landmarks, compute the shortest paths to other landmarks. Each landmark then computes the distances to other landmarks and estimates what we name the average hop distance size (AHDS) (cf. Figure 2.11). AHDS is the ratio between the average distance to the average hop count among all the other landmarks that this landmark is aware. This AHDS value is broadcasted into the network to represent the single estimated average size of one hop (in meters) for that landmark i. A wireless ad-hoc node with unknown position uses received AHDS value from closest neighboring landmarks to estimate the minimum distances to the landmarks. A minimum distance is a multiplication of a result from the hop count to a closest neighboring landmark i and then multiplied by AHDS value of that landmark i: $Hops\_from\_i * AHDSvalue$. Note that a wireless ad-hoc node receives only one AHDS update from a closest landmark. AHDS values are distributed in controlled flooding: once a wireless ad-hoc node gets this value from a closest landmark and forwards it, it drops all the subsequent ones. A location estimation of that wireless ad-hoc node is completed using the common trilateration method involving AHDS values and distances at least three landmarks.

The same approach is later used by Savarese *et al.* [89] with the name of Hop-TERRAIN [88]. Both algorithms use multilateration [89] as a positioning mechanism. But instead of APS through DV-hop, Hop-TERRAIN uses refinement of the node position estimation to ameliorate

the obtained position accuracy. The initial position of a node is estimated by Hop-TERRAIN or DV-hop algorithm. Then a refinement algorithm uses an iterative distributed algorithm in which the nodes update their positions in a number of steps. Note that, the value of the number of steps to refine a node position is not specified by the authors. During each step of the refinement algorithm a node broadcasts its estimated position and receives at the same time the positions and corresponding ranging (distances) estimates from its neighbors. The new position of the node is estimated by trilateration.

The refinement of Hop-TERRAIN algorithm converges after a number of iterations, and when the position update becomes smaller. However note that the refinement algorithm is also sensitive to ranging errors and does not take into account all inter-node distances. In fact, the refinement algorithm only ameliorates the estimated position over those reported by initial Hop-TERRAIN algorithm. Simulation results show that overall refinement algorithm can improve the accuracy of the estimated positions by Hop-TERRAIN by a factor of 3 to 5.

DV-distance propagation method on the other hand uses the same methodology as DV-hop with the difference that the cumulative traveling distance is used. The distance between neighboring nodes is measured using radio signal strength. Then, the estimated distance is propagated in meters rather than in hops. However, all distances hops won't have the same size like in DV-hop. DV-distance propagation method is also sensitive to measurement errors. Euclidean propagation method propagates the true euclidean distance to the landmark. The condition of such method is that a wireless ad-hoc node in order to be localized needs at least two neighbors wireless ad-hoc nodes having connectivity information from a landmark i. In addition, these two neighbors need to be neighbors each other or their distances should be known. The euclidean distance between the wireless ad-hoc node to localize and a landmark i is computed. This euclidean distance corresponds to the second diagonal of a quadrilateral formed among the node to localize, its two common neighbors and the landmark i.

"DV-based" algorithm and "Euclidean" are evaluated with NS-2 simulator. A random topology of 100 nodes is generated with an average node degree of 7.6 and diameter of 10. Nodes are placed in a random manner while the communication range, the connectivity as well as the density are the same throughout the network. Simulation results demonstrate a localization error of less than one hop from the true location when using APS with the three proposed propagation methods. In addition, with "DV-based" methods, almost all nodes are estimated, whereas an accumulation of "Euclidean" error produces some unreachable nodes. Moreover, simulations results using hop-Terrain algorithm [89] demonstrate that landmarks should be placed carefully at the edges in order to obtain a precise positioning algorithm. In fact, their results show that, largest errors occur along the edges of the networks.

Many studies have follow the DV-Hop method [90, 91, 92]. Recently, Zheng *et al.* [92] have proposed a range free localization algorithm in Wireless Sensor Network (i.e. WSN). Such range free algorithm uses a Neural Network Ensembles named LNNE for Localization with Neural Network Ensembles to improve the localization accuracy in WSN. The *LNNE*-based localization algorithm solely uses the diversity of the neural network components to build the sensor position prediction model based on the hop-count information (or connectivity information) of the wireless sensor network. The training Neural Network Ensembles gathers the hop-count and location information of all the landmarks in the network. The base station node broadcasts the information of the trained Neural Network Ensembles to all the sensor nodes in the network. That is the hop-count length of the shortest path between sensor node k and landmark i. Each regular node communicates with a landmark node through a multihop path. A regular node then estimates its location with the received hop-count information and the trained Neural Net-

work Ensembles as input to the Neural Network Ensembles. The performance of localization with *LNNE*-based localization is evaluated in simulation. *LNNE*-based localization is also compared with DV-Hop algorithm [29], centroid [92] and a single neural network-based localization algorithm.

During simulation these following factors are compared: the effect of the number of landmarks, the network density and coverage holes. Simulation results demonstrated a better performance with *LNNE*-based localization with 7 components Neural Network Ensembles in an area of size 50 m by 50 m. *LNNE*-based localization with 7 components Neural Network Ensembles achieved a performance of 3 m with 150 sensor nodes in total and 1.7 m with 400 sensor nodes. Instead, an average localization error ranging from 3.7 m (i.e. 150 sensor nodes) to 3.2 m (i.e. 400 sensor nodes) is observed with DV-hop [29]. Furthermore, a refinement of *LNNE*-based localization is realized to improve the performance of localization accuracy. The refinement part uses an enhanced mass-spring optimization algorithm (i.e. EMSO). EMSO algorithm uses the location information of both: the neighboring beacons and those of unknown nodes.

In EMSO algorithm, each sensor node is considered as a mass and all masses are connected each other with linear and elastic springs. A node adjusts its position using spring force computed from its neighbors. A node location is adjusted only if the new location reduces the local energy. The force on the sensor node k is pulled by a neighboring sensor node, which results in the total force on node k. The localization accuracy is improved by minimizing the total energy of the spring system. Minimizing the total energy of the spring system implicates to minimize the local energy of each unknown node. The algorithm converges if the maximum number of iterations T is reached. This refinement algorithm can be more advantageous than the one proposed in [89] as the spring approach can compensate measurement errors without propagating them throughout the grid [93, 94]. The refinement algorithm improves the localization error at about 24%. The algorithm achieves an average localization error ranging from 2.5 m to 1.2 m for respectively a network density of 150 and 400 sensor nodes and a landmark ratio of 0.2 in an area of size 50 m by 50 m. However, the Localization with Neural Network Ensembles reports an average localization error from 3 m to 1.7 m using the same simulation parameters.

The second category of range free algorithm exploits the signal strength measurements as a localization protocol. He *et al.* [85] present an Approximated Point-In-Triangulation based localization algorithm named APIT. *APIT* is an area-based range-free localization algorithm that compares the radio signal strength measurements. *APIT* localization approach first isolates the environment into triangular regions between landmarks nodes. A node may reside inside or outside an area if its presence is detected inside or outside of these triangular regions as shown in Figure 2.12.

As presented in Figure 2.12, if none of the neighbors of the blind mobile node are more closer to the three landmarks, the blind mobile node supposes to be in the interior of the triangle formed by the corresponding landmarks. This means that the blind mobile node has respectively measured from all the three landmarks, the strongest RSSI values compared to its neighbors. However, if one of the neighbor assumes being farther from landmarks than the blind mobile node, in this case, the mobile node is believed to be outside the triangle formed by the three landmarks. On the other words, the relative position of a node is based on the triangles. Then, the intersection (i.e. center of gravity) of all the triangles which may contain the node is performed by the node itself to estimate the diameter of the area in which it resides. Via extensive simulation, *APIT* is claimed to perform best when an irregular radio pattern and a random node placement are considered. Moreover *APIT* can make an incorrect decision (14% in the worst case) if the blind mobile node is near the edge of a triangle formed by a set of landmarks, and that some

Neighboring node 5

RSSI values measured by the neighboring node 5 from data packets sent by each landmark

	Measured RSSI values	Coordinates of each landmark	
LM_1	-65 dBm	3.5 m	8 m
LM_2	-88 dBm	2.8 m	4.2 m
LM_3	-94 dBm	5.8 m	7.72 m

Blind mobile node 1

RSSI values measured by the blind mobile node from data packets sent by each landmark

	Measured RSSI values	Coordinates of each landmark	
LM_1	-55 dBm	3.5 m	8 m
LM_2	-72 dBm	2.8 m	4.2 m
LM_3	-84 dBm	5.8 m	7.72 m

(a) Step 1 : collection of RSSI values from landmarks

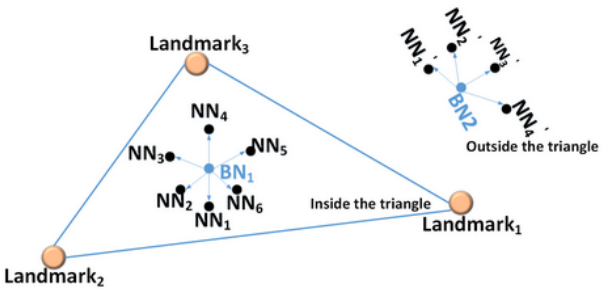
2- Blind mobile node is inside or outside a triangle formed among landmarks?

-Theorem: Received RSSI monotonically decreases with increasing distance.
=> If RSSIM1-BN > RSSIM1-NN5, BN is more close to M1 than NN5

-Lemma1: If the blind mobile node is inside the triangle $LM_1LM_2LM_3$, shifted the blind mobile node in any direction must lead to a new position nearer to at least one of the landmark LM_1,LM_2 or LM_3.

-Lemma2: If the blind mobile node is outside the triangle $LM_1LM_2LM_3$, when shifted the blind mobile node in any direction, there must be a direction in which the position of the blind mobile node is closer to all the three landmarks LM_1,LM_2 and LM_3.

Landmark (LM or reference node)
Blind mobile node
Neighboring nodes (NN)



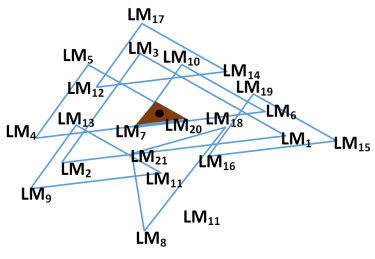
	Measured RSSI values by BN	Coordinates of each landmark		...	RSSI value measured by NN_5	...	RSSI value measured by NN_n
LM_1	-55 dBm	3.5 m	8 m		-65 dBm		-58 dBm
LM_2	-72 dBm	2.8 m	4.2 m		-88 dBm		82- dBm
LM_3	-84 dBm	5.8 m	7.72 m		-94 dBm		-96 dBm

Neighborhood network maintained at each node

(b) Step 2 : process that determines whether a blind node is inside or outside a triangle

3- Position estimation

-Repeat (2) with all possible audible landmarks combinaisons (aggregation)
-Determine the area with maximum overlap
-Calculate the center of gravity of this overlap area: the position of the blind mobile node



(c) Step 3 : position estimation

Figure 2.12: The area-based algorithm principle: *APIT* for Approximated Point-In-Triangulation

of its neighbors are outside that triangle farther from all its points. According to the authors, a blind mobile node can assume to be outside the triangle formed by the landmarks due to the effect of edge mentioned above. This is not the case in [85] as the radio model supposes an upper and lower bound on signal propagation. All nodes beyond the upper bound are out of the communication range while those within the lower bound are guaranteed to be within communication range.

APIT algorithm is evaluated in simulation with routing algorithms. The authors analyze how localization error can affect a variety of location-dependent applications i.e. target estimation, routing and object tracking. Tested routing algorithms are those whose routing decisions are based on knowledge of the geographic location. Simulation results of *APIT* algorithm demonstrate the delivery ratio of routing algorithm to increase with a small localization error and when changing the node density. With an average localization errors of 0.2 and 0.4 times the node radio range, the delivery ratios of the routing algorithms are very close to have no error. The authors have additionally demonstrated that some factors when increasing alone or together can decrease the localization error. That is the number of anchor nodes should be no more than 12 as well as the node density should remains under 6. However, the departure test of the *APIT* algorithm may not always be verified in an obstructed indoor environment. Particularly, a node to determine whether a neighboring node is closer to a given anchor uses the assumption that the received signal strength monotonically decreases in an environment. Such assumption may not hold in realistic indoor environments because of strong disturbances.

Instead, ROCRSSI [95, 86] originally proposed in [95] and extended in [86] is a famous range free based localization algorithm. It assumes a monotonic decreasing function of the distance, for the received signal strength loss measured between and emitter and a receiver. With ROCRSSI algorithm, a blind node localizes itself by comparing the received signal strength loss readings from several landmarks to those measured by each landmark from its neighborhood landmarks.

Let us denote by $RSSI(A, \mathcal{BN})$ a signal loss measured by a blind node $\mathcal{BN}$ from signal emitted by a landmark A. If $RSSI(A, B) \geq RSSI(A, \mathcal{BN})$, one concludes the blind node $\mathcal{BN}$ lies outside the ring with center A and radius $d(A, B)$; inversely, if $RSSI(A, \mathcal{BN}) \geq RSSI(A, C)$, then $\mathcal{BN}$ lies inside the ring with center A and radius (AC). Thus, $RSSI(C, D) \geq RSSI(C, A) \geq RSSI(A, \mathcal{BN})$, indicates that $\mathcal{BN}$ lies in the shaded ring with center C. To calculate any blind node position, the ROCRSSI algorithm generates a set of rings, each of them containing the blind sensor node. The algorithm then computes the intersection of the rings. The center of gravity of the intersection is the estimated position of the blind sensor node. The ROCRSSI algorithm pseudo-code is presented in Figure 2.13.

This distributed ROCRSSI algorithm is correctly tested with three Micaz sensor nodes in an outdoor testbed of size 40 m by 30 m. Testbed results report an average error of 5.18 m with 3 landmarks and indicate that anchors must be carefully placed for a better performance of ROCRSSI algorithm. ROCRSSI is also evaluated in simulation using various deployment strategies. Simulation results with a deployment of landmarks inside a terrain have better performance than deployment of landmarks at the edges of a terrain. In addition, this deployment should be uniform as the coverage of rings is reduced when landmarks are placed at the edges. In fact, a ring is cut off by the edges and at most only half of the ring is usable for the localization algorithm. Another conclusion from ROCRSSI simulation is a more number of landmarks i.e. more than 81 gives a better localization accuracy. A modified version of the ROCRSSI algorithm named ROCRSSI ++ is proposed by Frattini *et al.* [96]. ROCRSSI ++ takes into account the RSSI values of other unknown nodes, in addition to the one of the landmarks. Another improvement is derived from the ROCRSSI ++ algorithm: the power-Based ROCRSSI (PB-ROCRSSI) algo-

```
BN : Blind node
{L} : set of landmarks
{R} : set of rings (empty)

For all landmark A in {L}
        mingreather = max of float
        maxlower = 0
        For all another landmark B in {L}
                if (RSSI(A,BN) > RSSI(A,B) and RSSI(A,BN) < mingreather)
                        mingreather = RSSI(A,BN)
                        interRadius = dist(A,B)
                else if (RSSI(A,BN) < RSSI(A,B) and RSSI(A,BN) > maxlower)
                        maxlower = RSSI(A,BN)
                        externRadius = dist(A,B)
                end if
        end for
        And ring center A and radius  (interRadius,  externRadius) in {R}
End for
I = all space
For ring R in {R}
        I = I inter R
End for
BN is in R
```

Figure 2.13: The pseudo code of the ROCRSSI algorithm

rithm [87]. This algorithm uses data fusion. It combines multiple RSSI readings measured by varying the transmission power while keeping the same localization property of its predecessors. For instance, at 0 dBm, RSSI value achieves its maximum while at -15 dBm it achieves its minimum value. Based on ROCRSSI ++, PB-ROCRSSI algorithm exploits the increasing trend of the RSSI regarding the transmission power variation. The authors evaluate PB-ROCRSSI algorithm in a realistic testbed using sensor networks. ROCRSSI ++ is claimed to achieve a lower accuracy compared to ROCRSSI algorithm. ROCRSSI ++ exhibits a lower variance than PB-ROCRSSI algorithm. But, ROCRSSI ++ reports a localization error of more than 25% compared to ROCRSSI which has obtained a localization error of more than 38%. Instead PB-ROCRSSI shows a lower localization error than ROCRSSI ++. PB-ROCRSSI algorithm highlights a localization error of 12% which corresponds to a precision of 125 cm in a 31 squared meters environment with six landmarks.

Many published papers related to range free algorithms have revealed that most of the existing range free algorithms are evaluated in simulation. Even in the presence of a good radio channel, the acquired localization accuracies are coarse for most simulation-based range free algorithms. An inaccurate/roughly localization accuracy is sufficient for both outdoor applications or indoor applications that only rely on the fact that a blind node is inside or outside a closed environment like a room. However, when it comes to find the position of a person/object inside a room/-conference room or to track an elder person under medical supervision, some of the proposed algorithms can be immediately ruled out for indoor localization scenario. Accuracy requirements have encouraged the study of range-based localization algorithms. Such type of algorithms depend on the availability of absolute point-to-point distance estimates or angle information. Even though a more accurate position can be acquired with the use of ranging measurements between nodes, the accumulated error should be small, as it directly affects the localization algorithm.

2.2.4 Range-based algorithms

Ranging algorithms can be divided in two categories : those based on radio propagation model and those using empirical RF map based algorithms. Algorithms based on RF mapping uses the same techniques as fingerprinting described above. These algorithms match a specific location to a set of measurements (like RSSI readings) parameters in a given environment. Algorithms based on RF mapping include matching and classifiers approaches. Such algorithms consist of performing a measurement of RSSI values at various predefined locations of the area of interest. Such offline training phase permits to establish a prior radio map of RSSI readings. In online phase, when localizing an unknown target node, a matching is made between current collected RSSI values and the closest existing fingerprint using probabilistic approach or a deterministic approach (i.e. distance metric) [19, 17].

Deterministic techniques subdivide the physical environment into areas in which the positions are chosen to measure corresponding RSSI values in order to build a training dataset (i.e. in offline phase). A blind node location is estimated during the online phase using techniques that select the nearest signal strengths comparable to current RSSI readings. Thus, such techniques link RSSI readings to an area in the training set. An example of a deterministic technique is location-based K-Nearest Neighbor classifier [19, 18]. Probabilistic techniques however make usage of a probability distribution function to estimate the blind node location. Example of probability distribution function is the area with highest probability density [97]. In [97], the authors use particle filters methods in indoor wireless environment. Precisely, a recursive Bayesian filter is used to estimate the current location on a posterior distribution of the blind node location. The arbitrary a-priori distribution that computes a posterior distribution is based on the Monte-Carlo Sampling method. Even they are claimed to be more precise [97], such methods are also computationally intensive. These type of methods need to have a prior knowledge about the position of a node. Furthermore, if such techniques yield a relatively good accuracy, the calibration and the computation demanding, make them less popular for the use in dynamic indoor environment. Another drawback of these algorithms is that their accuracy depends on the accuracy of measurements. The ranging algorithms based on radio propagation model [98, 68, 99, 100] on the other hand do not involve a pre-deployment strategy like maintaining a signature database. Instead, these types of algorithms depend on the underlying channel and thus do not require detailed empirical measurements to generate a signal strength map and consequently have a low set up cost.

However as indoor environment is mostly subject to random variations due to obstacles, some research works use selective mode on RSSI data in order to obtain a better positioning precision. In [68], an accurate and self-adaptive RSSI-based indoor and outdoor localization algorithm is claimed. Such algorithm is building upon the log-normal shadowing model [24]. The proposed algorithm is named log-normal shadowing model with dynamic variance (*LNSM-DV*) and controls the dynamic variance. *LNSM-DV* hypothesis is based on the authors experimental results which state that the variation of the signal strength over distance is regularly described by the variance. In other words, one hundred groups of RSSI values are collected in each point over distances from 0 to 6.1 meters. Only twenty groups of data are selected to draw the curve of the signal strength in function of the distance for experimental results. The analysis of such number of experimental data has demonstrated the variance of signal strength to change regularly with distance. A Gaussian distribution is assumed. *LNSM-DV* expresses the log-normal shadowing model with a dynamic adjustment of the variance by proposing a polynomial function of 3rd degree (cf. [68]) between the variance of the RSSI and the distance. Linear Least Square regression analysis is used to estimate the parameters of the *LNSM-DV* model: a curve

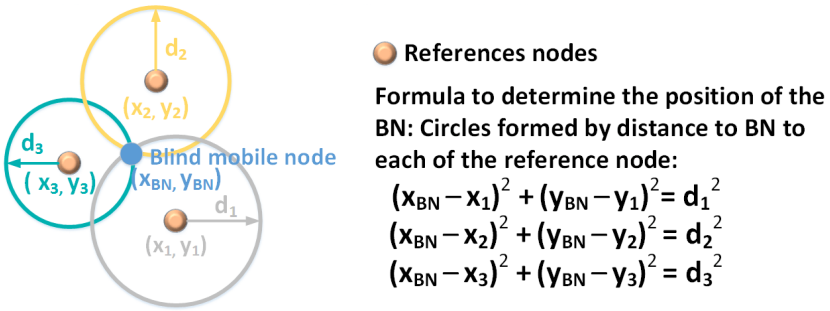


Figure 2.14: Trilateration principle

of the variance (of collected RSSI) in function of the distance is fitted to a polynomial function using an IBM predictive statistical analysis software named SPSS (i.e. Statistical Product and Service Solutions [101]). However, making a data selection in somehow controlled manner do not guarantee the repeatable property of the authors experience, neither the self adaptability of the proposed method in any indoor environment.

Range-based localization algorithms require the knowledge of ranging information between nodes. As mentioned the most famous ranging techniques include distance measurement, angle measurement. Distance can be evaluated using RSSI or time-based measurement techniques. RSSI evaluates the attenuation along the path [102]. Time-based techniques instead measure the propagation time of electromagnetic wave between an emitter and a receiver. Recall that atte-nuation is the loss of signal power in decibels (i.e. dB); It is a function of the distance between a transmitter and a receiver unit. The greater is the distance, the greater is the attenuation value and the weaker is the received radio signal. A full survey of measurement techniques as well as a comparison between various strategies accuracy are available in [16], [20]. Another survey and taxonomy is provided by Mao *et al.* [1]. Range-based localization algorithms require the knowledge of ranging information between nodes such as the measurement of distance or the angle.

Inter-nodes distance measurements techniques Localization based ranging uses distance information among wireless nodes to estimate the corresponding position. Distances at least three reference nodes are needed in most range-based localization algorithms such as trilateration. Recall that reference nodes usually called landmarks or anchor nodes are nodes with known positions. Assuming a Gaussian distribution and considering Free Space Path Loss prediction, the received power of the radio signal decreases as a function of the separation distance between a radio emitter and receiver. Knowing the distance separating a blind node and a landmark leads to a circle centered at the landmark and with radius the corresponding separation distance. Estimating the distance between the blind node and two landmarks convey to two circles. Distances from three landmarks lead to three circles. The intersection of these three circles provides the unknown position of the blind node estimated with trilateration. Trilateration is a geometric technique that utilizes the geometric property of reference nodes or landmarks locations to estimate the position of the blind node. At least three landmarks are required to uniquely position a blind node. An example of trilateration concept is presented in Figure 2.14.

From a theoretical point of view, trilateration leads to solve equations of circles. Figure 2.14 considers three reference nodes/landmarks named LM1, LM2, LM3 and one blind node named BN. We Assume the separation distance between the blind node BN and each reference node

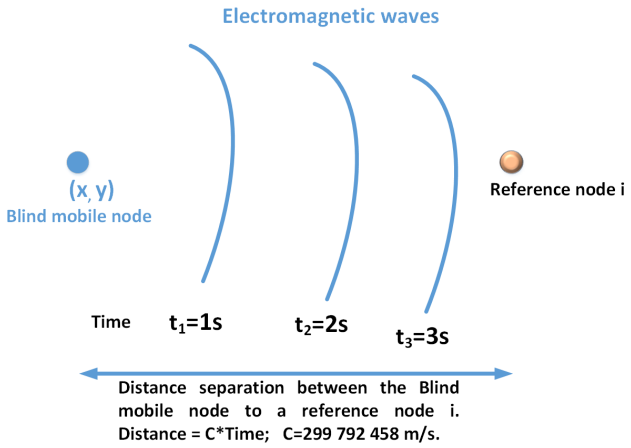


Figure 2.15: Time of Arrival measurement principle

is determined through RSSI or time-based measurements. Let us denote by d_1, d_2 and d_3 respectively the separation distance between the blind node and LM1, LM2 and LM3. d_1 represents the radius of a circle centered at the reference node LM1 and covering the position of the blind node. Having an information from a second distance between the blind node and let's say LM2 conducts to a second circle in which the blind node is belonged. Such second circle is centered at LM2 with radius d_2. The intersection of these two circles results in two distinct points in $2D$ positioning while a circle is obtained in the case of a $3D$ positioning. In a $3D$ positioning, the knowledge of one reference point leads to a sphere centered at the corresponding reference point. Theoretically, a presence of a third circle that associates the position of the blind node leads either to a unique intersecting point (i.e. the position of the blind node in a $2D$ positioning) or to two intersecting points in a $3D$ positioning. Thus, theoretically, a fourth reference node is required to position a blind mobile node in a $3D$ localization system.

The RSSI [102] is measured through the radio frequency communication interface from a signal emitted by a neighboring node. The RSSI or signal strength measurement uses a propagation model to describe the path loss attenuation with distance. Time-based techniques can be evaluated on a receiver node with the use of the one way time of flight [103], [104] of the signal emitted by another node. Such measurement is usually known as Time of Arrival (i.e. ToA) of the signal. Figure 2.15 shows an example of ToA measurement.

Time measurement techniques include several variants along with the Time Difference of Arrival (i.e. TDoA) and the Roundtrip Time Of Flight (RTOF). TDoA [105, 106] uses the same measurement technique as ToA. It is based on a reference ToA and then measures the difference of the time during which the signal travels a point (e.g. A) to another point (e.g. B) as shown in Figure 2.16. For instance, with TDoA-based localization, a blind node may send a signal that is received by several wireless nodes. The blind node can use a reference ToA based on the time by which the emitted signal is received by one receiver node (e.g. landmark). Then, the blind node can compute its position based on the time difference of the received signal among different receivers. For TDoA-based localization, a synchronization with the blind node is not required. However, all landmarks should use the same frequency channel and communicate with the same blind node [106, 107, 108].

Instead, RTOF [20], [109] estimates a full return propagation time, following the response of a receiver, for instance, at a radar signal's query. Performing time measurement is not obvious in distributed network context, as a clock synchronization between each radio emitter-receiver

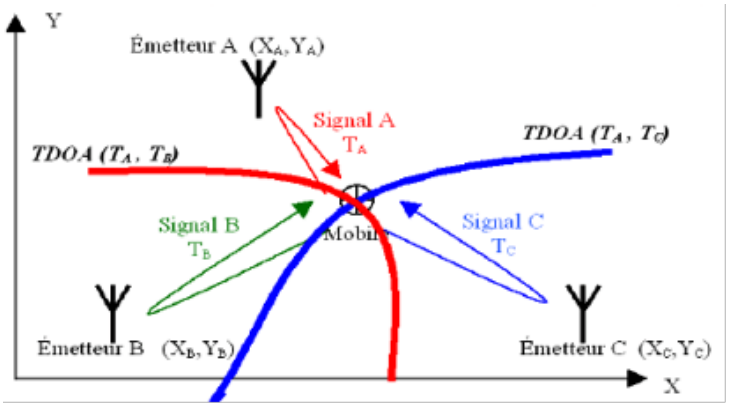


Figure 2.16: Two Time Difference of Arrival measurement principle [105]

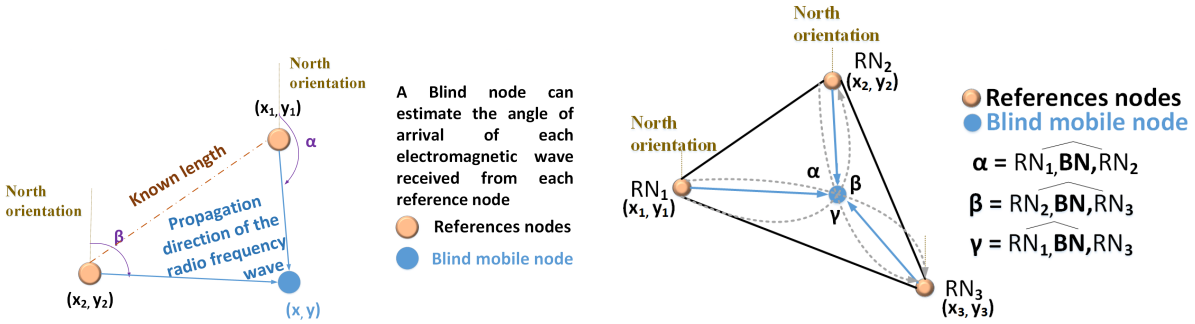
is required when using the ToA technique. Moreover localization-based ToA offers high levels of accuracy while at the same time RTOF requires relatively fast processing capabilities to resolve timing differences for fine-grained measurements. Particularly, over short distances time measurement requires a more faster processing for timing differences resolution which can make them a bad candidate for wireless ad-hoc sensing network [89]. No synchronization is required with RTOF since time measurements are made on a two way return from an emitter, the same clock is used twice. However, according to works in [20], it is currently difficult to know the exact delay/treatment time required by the receiver to send the signal, which is an additional source of errors. Instead, for very large or medium systems, the delay can be ignored if it is small compared to the transmission time which is not the case for small systems. Hence RTOF metric requires a real-time and an accurate system.

Angle measurement techniques Angle measurement techniques measure the Angle of Arrival (AoA) of a received signal at different nodes. From a geometrical point of view, estimating the position of a point inside a triangle using angle of arrival technique requires to know the position of the vertices of the triangle as well as the angles by which the interior point sees these vertices. This so-called triangulation [110] is presented in Figure 2.17.

AoA scheme [111, 110, 112, 113, 114] uses an antenna array and a directive antenna to estimate the direction of arrival of a signal. Location-based AoA as presented in Figure 2.17 is somewhat similar to the principle of trilateration used by the GPS system. The main difference between both techniques relies on the fact that with triangulation, a blind node (defined by the interior point in the principle of triangulation) knows the angles between the sides of the triangle instead of distances to the vertices.

In Figures 2.17(a) and (b), knowing the distances $\mathcal{BNRN}_2$, $\mathcal{BNRN}_1$, $\mathcal{BNRN}_3$ and the positions of the three reference nodes $\mathcal{RN}_1$, $\mathcal{RN}_2$, $\mathcal{RN}_3$, trilateration can be used by the blind node $\mathcal{BN}$ to infer its position.

Instead, if the blind node estimates all angles it forms with the three reference nodes: $\mathcal{RN}_1\mathcal{BNRN}_2$, $\mathcal{RN}_2\mathcal{BNRN}_3$, $\mathcal{RN}_1\mathcal{BNRN}_3$ and having information of the coordinates of the reference nodes, position can be estimated using triangulation. Triangulation [110] uses trigonometrical approach to calculate the position of the blind mobile node determined by the reference points ($\mathcal{RN}_2$, $\mathcal{RN}_1$, $\mathcal{RN}_3$) and known angles. Remark that distance can be acquired from several AoAs. In general, two dimensional triangulation requires two angle measurements and one length measurement like the distance between the reference points. In three dimensions,



(a) Angle of arrival measurement principle

(b) Example of position calculation ($D(X1, Y1)$) with angles measurement from reference points [110]

Figure 2.17: Example of position calculation with angles measurement

one length, one azimuth, and two angle measurements are needed to estimate a precise location. The azimuth angle is defined as the horizontal angle of the observer's bearing respective to a reference dire-ction: usually the northern point of the horizon directly beneath a concerned object.

Note that, AoA and time-based measurements lead to a more accurate ranging estimation than those based on RSSI measurements. However, the accuracy is achieved at the expense of higher equipment cost. Angles measurement require the use of directed or phased antenna array. This approach relies on the use of several directional antennas and therefore given the differences in arrival times and the geometry of the receiving array, it is then possible to compute the angle from which the signal is emitted. Besides the need to acquire an extra hardware, maintaining the antenna array can be prohibitive in size and hardware costly. Moreover, the required clock synchronization of ToA technique counts as an extra computation step in the localization process, and can degrade the localization accuracy [115]. RSSI is therefore the only simple available and feasible measurement on embedded devices; but it has a low accuracy compared to ToA and AoA-based techniques [116].

Intuitively, a precision of location estimation can be enhanced by combining two or several measurements. Classes of hybrid schemes have been proposed to achieve a more accurate location estimation associating either ToA and AoA [117] or AoA and RSSI [118] or even TDoA and RSSI [3]. Works from Chen *et al.* [117] integrate all the available heterogeneous measurements by combining both: ToA estimations at seven base stations and AoA information at the serving base station to provide location estimation of the mobile station. In addition, non line-of-sight (i.e. NLOS) errors are eliminated using all possible intersecting circles obtained from ToA and AoA measurements in order to locate the blind mobile node. Such method does not involve any a priori information about the NLOS error. Jiang *et al.* [118] have introduce a hybrid localization algorithm that fuses RSSI differences with AoA. The proposed hybrid localization algorithm achieves a localization errors of about 124 centimeters using 81 tested grid points in a 10 m by 10 m indoor testbed with two landmark nodes. They estimate AoA by comparing the RSSI values of landmarks signals received from two perpendicular-oriented directional antennas installed at the same place. With the proposed hybrid positioning algorithm named ALRD (i.e. AoA localization with RSSI differences of directional antennas), a sensor node can localize itself within 0.1 second by observing the RSSI values from signals emitted by two landmarks. However, this method is not immune to measurement errors because the RSSI is susceptible to interference from the environment and from the external factors (i.e. misplaced objects relative

to the radio communication path, moving people). Consequently, the estimated location may deviate from the real location.

In [3], a hybrid TDoA and RSSI extended kalman filter-based localization algorithm without a motion model is proposed and evaluated in a warehouse of size 40 m by 60 m during the working time. The authors Experimental approach use eight landmarks placed at a height of 2 m in order to track a man. A man holds an IEEE 802.15.4 wireless node on his hand. The proposed hybrid localization algorithm achieves an accuracy of up to 3 m in area of the warehouse with only one or two landmarks with line of sight. This accuracy is below 1 m in other parts of the warehouse. TDoA measurements is used as a primary source of location information. A blind node broadcast a signal that is used by the landmarks to estimates the time differences. Then RSSI-based radio propagation measurements are used to correct the estimation. The algorithm only considers "good" RSSI values coming from a minimum distance: the smallest distance between the blind node and a landmark defined by TDoA measurements. More precisely, measured RSSI values are considered if they are superior to the one computed based on TDoA measurements. Then position is estimated using multilateration and the measured time differences.

2.3 Summary and discussion

The unpredictable nature of wireless medium is challenging for radio communication, especially in indoor. Consequently, over these ten years several research efforts are made on indoor wireless localization. Besides, advanced development in numerous communicating low-power and low-cost devices through the Internet of Things (IoT) coupled with the spatial localization of elements belonging to a wireless sensor network play an important role in identifying the origin of the data measured by a sensor node. Obviously these data lose their meaning if no location information is associated. In addition, several applications using localization techniques are envisioned in various domains related to the Internet of Things such as personal health monitoring, objects and persons tracking, intelligent homes and cities or environment study and preservation. However, before employing localization information in the Internet of Things applications, the following question should be resolved: how to obtain a precise distance information from unsteady and obstructed indoor environment?

For the sake of simplicity and the ease to use measurement technique, the RSSI is preferred to localize a wireless sensor node instead of measurements based on time of arrival or on the direction of incoming signal (AoA). Time-based measurements require a clock synchronization among all emitters-receivers as well as identifying the LOS in multipath environment. One TDoA measurement includes two times measurements from two different signals of two different landmarks and clock synchronization among landmarks. This makes TDoA difficult when it comes to identifying the outlying landmark in case of multipath [3]. Instead, angle of arrival measurements require the use of an antenna array, which can be prohibitive in size and thus power consuming.

We have presented several localization algorithms based on these two categories: range-based [19],
[116], [17], [98], [102], [68], [99], [100], [68], [5], [118] and range free algorithms [29, 89, 28, 88, 85, 31, 83, 90, 86, 91, 84, 92]. Most of these localization algorithms are distributed based on how the position is estimated: that is each wireless sensor node computes its position itself based upon received data from its neighboring sensor nodes. A Localization algorithm is also considered as distributed if a wireless sensor node can compute its position and those of its neighbors. Such

algorithm can be additionally fully distributed if sensor nodes are capable of communicating with direct neighbors as well as with intermediate neighbors through connectivity information. Note that wireless sensor nodes are randomly deployed across the environment and thus, connectivity computed either via number of hops or proximity, informs how close a neighboring node is. However, the distance measurements should not involve a lot of communication messages among nodes in order to avoid battery consumption while saving the energy. Consequently, a localization algorithm should allow interference by tolerating/authorizing messages losses. It should also permit refinement using information from neighbors to improve the node positioning accuracy [89, 92]. Such classes of algorithms are more fault tolerant and self organizing compared to centralized one. In centralized algorithm, the location of the entire sensor node in the network is performed by a single node and conducts to a single point of failure.

An excerpt from a literature review shows that several contributions studied and preferred range-free techniques [29, 88, 85, 119, 86, 28, 31, 83, 90, 86, 91, 84, 92]. However, such localization mechanisms do not provide sufficient and fined-grained (i.e. centimeter level) precision to localize a mobile node in a room context (i.e. office, classroom and so on). RSSI based ranging algorithms instead require inter-nodes distances estimation. Such models can be more accurate but are most affected by ranging errors depending on the environment characteristics. Range-based localization algorithms use two types of mechanisms. Those based on empirical measurements and those using a radio propagation model in order to map measured RSSI values to corresponding distances. However, the use of radio propagation models are preferable to empirical fingerprinting method. The latter requires extensive calibration and maintaining a signature database (i.e. RSSI to distance). Such fingerprinting models also suffer from the database update recurrence each time the environment of interest changes: presence of a human being, changing the location of an object and so on. Radio propagation models are of two types: empirical and theoretical. Empirical radio propagation model are preferred as they are much simpler to estimate. Such models use statistical properties to describe the radio propagation over parameters based on measurement data. Instead, theoretical or deterministic models are more accurate, but are site specific and require geometry information about the environment. Theoretical models employ analytical model (e.g. Maxwell's equations) to estimate the propagation of radio wave. Theoretical models are more accurate then empirical one [49, 24, 47] but require more computation. Several radio propagation mechanisms are presented in Figure 2.18 from free space; log-normal shadowing; partition loss and wall attenuation factors.

However, among the presented propagation models, the log-normal shadowing seems a good compromise between complexity and simplicity as well as feasibility. This model captures more environment characteristics via few parameters and is not much complicated to calculate compared to free space model. In addition, the number of parameters to estimate with log-normal shadowing model is less compared to modified models for wall or partition losses propagation. The log-normal shadowing can be applied to indoor and outdoor environment. Also the model parameters can be easily implemented on wireless sensor nodes. Therefore, the log-normal shadowing model is more pertinent for the use in RSSI-based localization algorithm in order to be applied in confined area like room, office environment.

Several RSSI-based indoor wireless localization methods applied on wireless sensor nodes have been used in many proposed localization algorithms (cf. Table A.1). However, none [116, 120, 121, 122, 5, 6] or few of them [98, 102, 123, 99, 7] allow the under two meters or centimeter-level of accuracy within buildings. Conversely, such type of sub-range is necessary in smart home context for example. Existing range-based localization approaches are classified in two categories: one-hop and multihop localization. One-hop ranging localization techniques use direct ranging measurements from neighboring nodes to estimate the position. One-hop localization technique

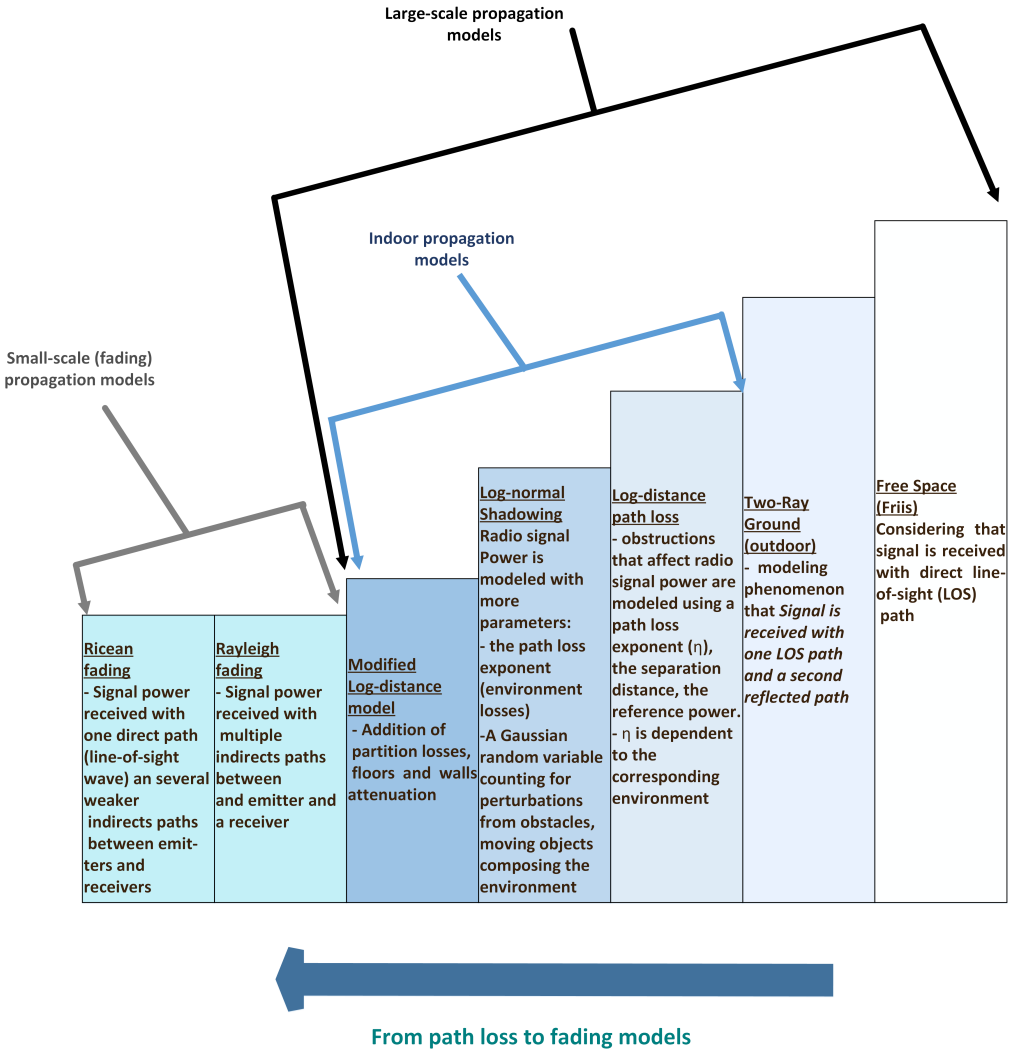


Figure 2.18: Summary of presented propagation models: from free space - shadowing - fading

can be a reliable localization technique if direct ranging information are accurately estimated. In fact, additional sources of errors can be mitigated in one hop localization mechanism as ranging measurements are not propagated through the network but directly exploited for position estimation.

Furthermore, in a wireless sensor network architecture a blind node to localize can be within a number of hop to the landmarks, with $Number\_hop \geq 1$. Consequently multihop ranging techniques work in cooperative way either by exploiting the number of hops or by propagating distances through the network or even using information from the unknown nodes to enhance the location estimate. This type of localization is also known as distributed [124, 93, 7], cooperative [125], GPS-free [28] or GPS-free-free [31]. With this type of algorithm, in most cases, multiples sensor nodes positions are estimated at the same time together and connectivity or ranging information from unknown sensor nodes can participate to the localization mechanism estimate. Moreover, from our opinion, another way of considering multihop localization mechanism is formalized in the following: regular nodes that are of one-hop neighbors from the landmarks can collect the positions information of the landmarks as well as the connectivity information to estimate their positions through a one-hop localization technique. Then, regular nodes which are at least of two-hops neighbors of the landmarks can estimate their positions in distributed manner based on the position of the one-hop neighbors of the landmarks (additional newly available information) in addition to the landmarks positions.

Indeed one-hop neighboring nodes can broadcast their newly estimated positions and those of their neighboring landmarks to their neighboring regular nodes in downlink (i.e. nodes that are of two-hops from the landmark nodes). These two-hops neighbors estimate their positions with the same one-hop localization algorithm and the new information available from their one-hop neighbors (uplink to the landmarks). This mechanism continues until all the sensor nodes locations are computed. These two-hops neighbors from the landmarks can then broadcast their newly estimated positions to their neighbors (3-hop neighbors from the landmarks) and those from one-hop neighbors of the landmarks. The advantage with this method is that the same one-hop localization algorithm is used to estimate the position of the nodes lying in more than one hop from the landmarks. In addition, local broadcast is used for communication with immediate neighbors. The broadcasted messages are then limited to one-hop neighbors within the communication area in the wireless sensor network with low range and low transmission power. Thus, the accuracy of the position estimate depends on the accuracy of the localization mechanism. Note that one-hop localization technique can be a reliable localization technique if direct ranging information are accurately estimated. In this case, additional sources of errors can be mitigated as ranging informations are not propagated through the network but directly exploited for position estimation.

Afterwards, when the network converges, that is all nodes have estimated their 2D positioning coordinates, such information can be then exchanged among nodes using controlled flooding: each wireless sensor node in the network sends its estimated $2D$ coordinates to its neighbor nodes that have not yet received such estimated positions to avoid sending too many messages unnecessarily. Therefore every node will have a map of the $2D$ positions of the entire nodes in the network. This network $2D$ positioning map can then be updated at each new insertion/arrival or departure of wireless sensor nodes. Besides, if the position of the one-hop localization algorithms are affected by ranging errors, such errors will also affect the location estimate of the other hop neighbors which are relative to the one-hop neighbors positions.

Recently, in [7], a collaborative/cooperative spring/force based localization algorithm is proposed. The spring algorithm is previously introduced in [94] with the log-normal shadowing

propagation model. The spring algorithm can be viewed as a network of nodes connected by springs (using the concept of graph in practice) as presented in Figure B.1 in appendix B. However, an accuracy of under 2 m is difficult to achieve in indoor localization for instance in an indoor environment of about 11 m by 9 m [7, 94]. Such achieved localization accuracy can not make a distinction among several users locations inside a room like a conference room for example. In addition, from the authors results in simulation [94] and in experimentation [7], increasing the number of sensor nodes with the spring localization mechanism increases the network connectivity i.e. the number of neighbors of a sensor node and degrades the localization accuracy [126].

Therefore, given the localization results of recent research in indoor environment, obtaining a precise fine grained and easy-to-use RSSI-based localization algorithm in indoor remain a challenge and an open problem. Indoor environment is a difficult context to implement an accurate localization algorithm as the radio signal has to deal with frequent changes. In indoor wireless mobile communication system, electro-magnetic wave propagation is not only altered by reflection, scattering, diffraction and signal fading but those factors can lead to a dynamic variation of signal strength over time and distance due to locals environment impediments. Such effects can also be induced by external factors like the presence of an obstacle in the middle of the communication link disturbing at the same time the direct path between an emitter and a radio receiver. Such external factors may cause shadowing and signal fading. Thus an understanding of the indoor propagation effects is needful to design a suitable communication system and also a precise and accurate localization algorithm. Poor ranging measurements or a bad estimate of the inter-sensor nodes separation distance will have an impact on the estimated position.

Moreover, note that most of the proposed models are tested in simulation with NS-2 network simulator, OMNet, WSim/WSNet. However in most cases the accuracy is dependent to the simulation environment that is the used simulation parameters. In reality, obstacles are not considered in simulation environment. Simulation can be used to test the performance of the implemented protocol as well as the scalability, however such protocols should be evaluated in realistic environment for real accuracy validation. In [127] which follows the works in [128] the credibility and the quality of wireless network simulation is discussed. The authors come up with the conclusion that an arising question about the simulation critique is the inaccurate explanation of the used simulation setup. Furthermore with the existence of large scale wireless sensor nodes experimental platforms (e.g. 1000 nodes in total on four different platforms), protocols are easily tested for scalability reasons as well as for evaluation in realistic environment. Even though the energy consumption is not targeted, any algorithm to be used in sensor network should be designed by considering a cost-effective and accurate system with less energy consumption. These elements are addressed in Chapter 4 in our communication protocol. In addition, depending on the type of application (e.g. localize an elderly in a home, babies in a nursery), a fine grained accuracy can be demanded which prevent us from using range-free algorithms.

In the following, we first perform several indoor measurements campaigns to study the behavior of the RSSI under realistic indoor environments (uncontrolled propagation behavior). Such measurements campaigns realized in dynamic real life indoor environments guide our research for the development of accurate indoor localization algorithms applicable to sensor networks. Our goal is not to have a controlled environment or to select the measurement samples but we aim to implement indoor localization algorithms with the only use of the RSSI metric by analyzing and targeting the physical relationship between the RSSI and distance. Thorough experiments presented in Chapter 3 help to better characterize the indoor wireless channel as well as to identify discrepancies/deviations that affect the performance of the realistic indoor

wireless channel. We also validate in Chapter 3 the usage of the log-normal shadowing model in indoor context with presence of furniture, people moving inside the environment, wireless access points, wood and glasses. Using an alternate log-normal shadowing model, we introduce in Chapter 4 three one-hop RSSI-based localization algorithms to enhance indoor positioning accuracy using IEEE 802.15.4 radio chips in 2.4 GhZ band. We finally propose in Chapter 5 a fully distributed localization algorithm in wireless sensor networks. Proposed RSSI-based indoor positioning algorithms are adaptable to any indoor environment without either no need to prior expensive calibration as performed in the fingerprinting like localization algorithm or no maintenance of a signature database. We additionally show in the same Chapters 4 and 5 that our proposed RSSI-based one-hop and distributed localization algorithms can achieve meter and even centimeter-level of accuracy without no extra hardware.

Chapter 3

Indoor measurement campaign: results on experimental measurements of the Received Signal Strength Indicator

Contents

3.1 Introduction

As studied in Chapter 2, preferred empirical propagation models are based on approximations in order to be general. Recall that empirical propagation models are tuned with measured data and use statistical properties from recognized features during experimental studies to describe the radio wave propagation over few parameters. Among existing empirical models [22], [37], the log-normal shadowing propagation model is the simplest to estimate and a good compromise when considering feasibility and complexity. Not many parameters are used to determine any indoor or outdoor propagation behavior. However, as we mention in Chapter 2, indoor wireless mobile communication system leads to dynamic variations of signal strength over time, distance due to local environment impediments and reflection from obstacles. As our localization algorithms aim to be adaptable to any environment characteristics as well as performed under uncontrolled propagation behavior, we assess in this chapter the extent of the approximations under which the

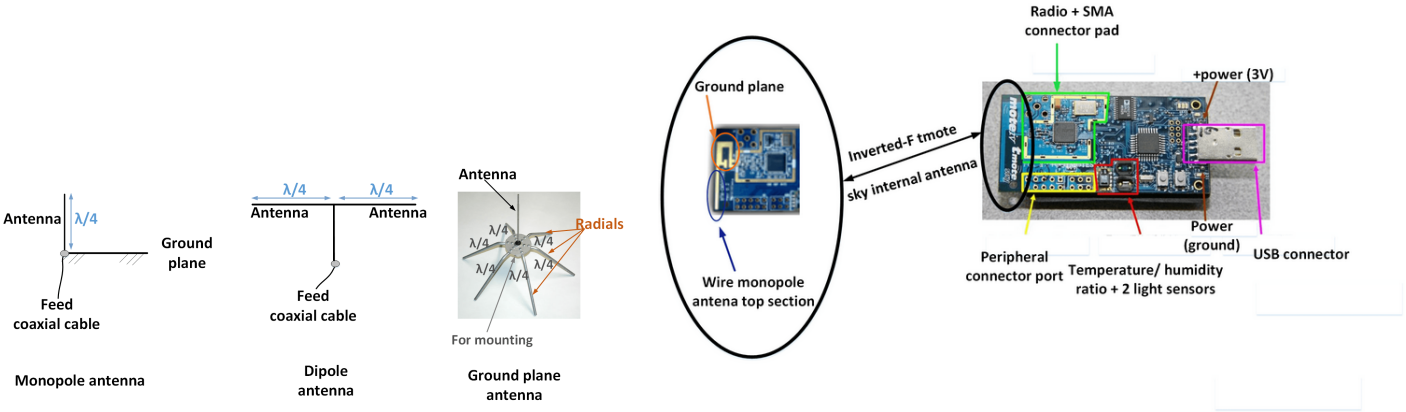
log-normal shadowing propagation model is based. We also characterize the equipment we use: Moteiv TMote Sky nodes. We identify their behavior in indoor environment, especially to get an idea of the variation of their characteristics that is never mentioned in the datasheet. Such characterization should allow us to better interpret the obtained results. For this, we conducted a series of measurements with multiple nodes in different environments.

For a better understanding of the experimental measurement campaign results we describe the type of existing antenna configuration and the one used by a Moteiv TMote Sky node. An antenna is a specialized transducer that converts the radio frequency (RF) power into electric and magnetic fields or vice-versa. An antenna is required for each wireless device. A radio wave is sent and received from an antenna also known as mechanical conductor device. A signal radiating from an emitter and a receiver antenna is transmitted usually by a transmission line for instance, a coaxial cable. There are many types of antennas. However, most of them are from the dipole configuration or the ground plane antenna[1]. Electromagnetic waves are made from inseparable electric and magnetic components. An electromagnetic wave is characterized by a frequency and a wavelength. A radio signal is a form of electromagnetic wave radiation. A dipole antenna is a $\frac{1}{2}$ wavelength structure and straight electrical conductor. From end to end, it is divided into two equal $\frac{1}{4}$ wavelengths linked at the center with a transmission line. A monopole antenna is one half of a dipole antenna and mounted above electric conductor called ground plane. This one half retained is normal to the ground plane. If the ground plane is large enough, the half-dipole behaves exactly like a whole dipole because its reflection in the ground plane will replace the missing part [129, 130]. A Tmote Sky mote uses an internal antenna of $\lambda = \frac{1}{4}$ monopole wavelength. As said in Tmote Sky datasheet [131] and shown in Figure3.1(b), this antenna is Inverted-F (i.e. I-FA) wire monopole where the top section is folded down to be parallel with the ground plane. The Tmote Sky internal antenna effective transmission range is about 50 m in indoor.

Printed circuit board (i.e. PCB) antenna are used to minimize the manufacturing cost while simplifying the design of sensor nodes [130]. A ground plane on a PCB also generally known as a terminal of the power supply consists of copper and is connected to the sensor node circuit's main radiating element. It is usually employed as a return path for electricity from many different components. Tmote sky and Telos are designed with PCB antennas. However, in [130], the authors stated that the performance of the PCB can be poor due to dielectric loss caused by poor circuit board material and interference by coupling with other lossy components and circuit board. Tmote Sky and Telos are designed with PCB antennas. It is better to use external antenna which is isolated from noise induced by the circuit board. The radiation pattern of monopole antennas above a ground plane are also known from the dipole result. The only change that needs to be noted is that the impedance of a monopole antenna is one half of that of a full dipole antenna. The directivity of a monopole antenna is directly related to that of a dipole antenna. Ground plane antenna is a vertical antenna. A ground-plane antenna is a variant of the dipole antenna presented in Figure 3.1(a): one half of a dipole mounted vertically. Its lower section consists of straight elements called radials. There are two or more radials, each measuring $\frac{1}{4}$ wavelength. It is composed of coaxial cable connected to main radiating element or earth ground. The shield below the antenna side connects to the radio. The ground plane antenna is for ground wave propagation. The way the ground plane antenna is made (i.e. with radials), it should have an omnidirectional pattern.

An understanding of the large scale propagation effects is very important to design a suitable communication system. Thus, characterizing the mote we use in our testbed helps to better

[1] `http://electronicdesign.com/wireless`



(a) An overview of a ground plane [132], monopole and dipole antennas

(b) An overview of the Tmote Sky wire monopole antenna

Figure 3.1: An overview of types of antennas

understand RSSI-based localization accuracy results. In the following we have evaluated the antenna gain and tested the omnidirectional pattern of the Tmote Sky node [131]. We measure, in an anechoic chamber, the signal attenuation obtained at a constant distance between an emitter and a receiver in various orientations. To determine the relative positions of the Moteiv Tmote Sky sensor nodes (i.e. Emitter-receiver) that are favorables or unfavorables to the balance of the radio link, we have conducted measurements of RSSI values between an radio emitter and a radio receiver in an anechoic chamber situated at Telecom ParisTech. Such measurements are performed to more characterize the radiation pattern of Tmote Sky motes. However, as shown in Figure 3.1 we should notice that the I-FA antenna is composed of two parts: a vertical part that radiates and an horizontal part linking to a power via coaxial cable (including a ground plane, battery and spring). Hence, the component with metal circuit may have an effect on the radiation pattern of the Tmote Sky node diagram. Furthermore, works from Lymberopoulos *et al.* [133] clearly advocate that different antenna orientations with respect to orientation angles (i.e. 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315 °) can yield to different sets of RSSI values for the same distances between an emitter and a receiver. Starting from these results, it is desired to characterize the antenna of the wireless sensor node that is involved in our study. A question arises: does Tmote Sky node have an on-board directive antenna?

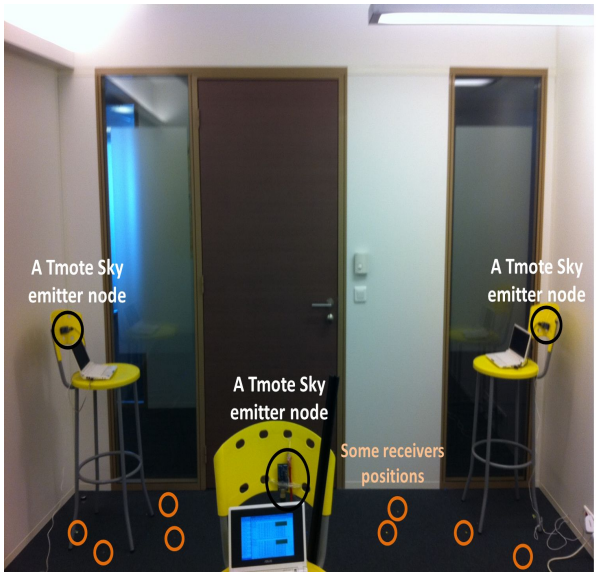
As localization performance depends on the environment and the used model to infer distance from path loss, we also evaluated through several experiments the RF signal propagation over few different indoor testbeds with various setup and sizes. We also investigated whether the log normal shadowing propagation model, often used to model indoor propagation, is enough accurate to capture irregularities in an office environment. Our experimental works are preceded by the results of Lymberopoulos *et al.* [133], Stoyanova *et al.* [134] (both are applied on outdoor environment) and of Wang *et al.* [135] while demonstrating in indoor environments, the variations observed in the measured RSSI values do not depend on a directional antenna. In the following, we do in-depth experiments in indoor environments with different configurations to better quantify these RSSI fluctuations. Our aim is to reduce error in the distance measurement, an element governing an accurate range-based indoor localization algorithm. Deep analyses and a quantification of RSSI fluctuations can further prevent the use of a complex localization algorithm that can be time and power consuming regarding wireless sensor network properties.

Notice that, our aim is not to impose a particular direction to a receiver node, in other terms our goal is not to constrain any person/object carrying a sensor node and wishing to accurately localize itself within a room/confined area to adopt a certain direction. Instead, our objective is to minimize the interference effects (less variations on measured RSSI values) while proposing accurate, precise and easy to use localization algorithms. Such localization algorithms should be adaptive to any wireless indoor environment dynamics without no need for calibration, nor fingerprinting as mostly did by lots of algorithms proposed so far. We do not also want to deal with complex localization algorithms in order to pay attention to the wireless sensor network (i.e. WSN) energy saving requirements.

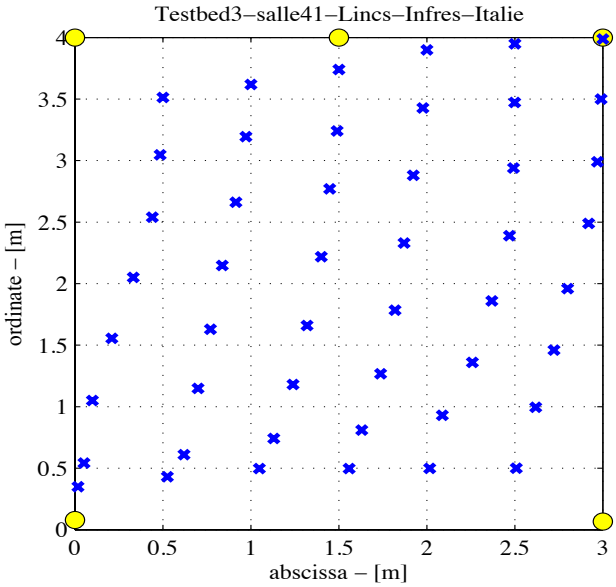
3.2 Comparison of the most representative contributions on RSSI-based localization algorithms

A good relationship between RSSI and distance is possible in controlled ways by exploiting certain factors as in [68]. We recall that in such works, a selection of RSSI values is made from a measured data samples. Moreover, works in [123] have shown that we can have a good relationship between the measured RSSI values and the corresponding distance by controlling the radio frequency, the number of the reference measurements, a specific node orientation. Several other works study the threshold beyond which the RSSI readings from the received signal are considered to be incoherent. However, a small disruption in the physical environment is sufficient to bring the measured RSSI values for a given distance out of the expected range. Given a link connecting a landmark and a receiving target node, the level of the received signal strength fluctuates considerably as a function of the position of a blind mobile node in an environment. Such phenomenon is due to objects blocking the line of sight path (i.e. shadowing phenomenon). Works from Elnahrawy *et al.* [116], Bahl *et al.* [19], [5], [7] have shown that the RSSI metric is not suitable for indoor localization due to the above mentioned propagation effects. [136] also demonstrated that RSSI approach performs poorly in indoor environment but did not look out why this phenomenon occurs. Conversely, Srinivasan *et al.* [137] mention that RSSI is not such a bad indicator when its value is above a certain threshold, measured at -87 dBm on the authors' experimental platform composed of IEEE 802.15.4 MicaZ nodes. While the works of Srinivasan *et al.* [137] provides good results, they do not provide much details on the testbed. Their conclusion may not hold in a realistic environment.

The system described by Hara *et al.* in [98] shows by simulation that the probability distribution factor (i.e. PDF) of the received signal power follows an exponential distribution with a power decay factor. The larger that power decay factor is, the more accurate the estimation of location should be. The authors also demonstrated that the number of RSSI measurements can improve the location estimation performance more than the number of sensor nodes. The same PDF is used as an indoor propagation model by Zemek *et al.* in [138]. The authors propose to reduce the traffic load ratio by studying a threshold beyond which RSSI values should not be transmitted to a given sink node for data aggregation. While this approach reduces the traffic load to the sink, it can still affect the target location estimation accuracy and even have a poor estimation of the link quality. Therefore several contributions show that the accuracy that one can expect with RSSI measures is very limited. A classification of the most representative ad-hoc and 802.15.4 wireless sensor networks localization algorithms in indoor environments with the use of RSSI metric is presented in Tables A.1 and A.2 in appendix A. Such classification concerning most of the state-of-the-art algorithms gives insight on achieved location accuracy against the number of used landmarks, usually $Nb_Landmarks \geq 10$.



(a) An overview of semi-furnished office testbed (#1) of size 4 m by 3 m)



(b) Testbed 1: (5 landmarks, 48 receiver node unknown positions)

Figure 3.2: Locations of the landmarks (yellow bullets) and of the test positions (blue stars) in testbed (#1)

3.3 Experimental conditions: testbeds and hardware description

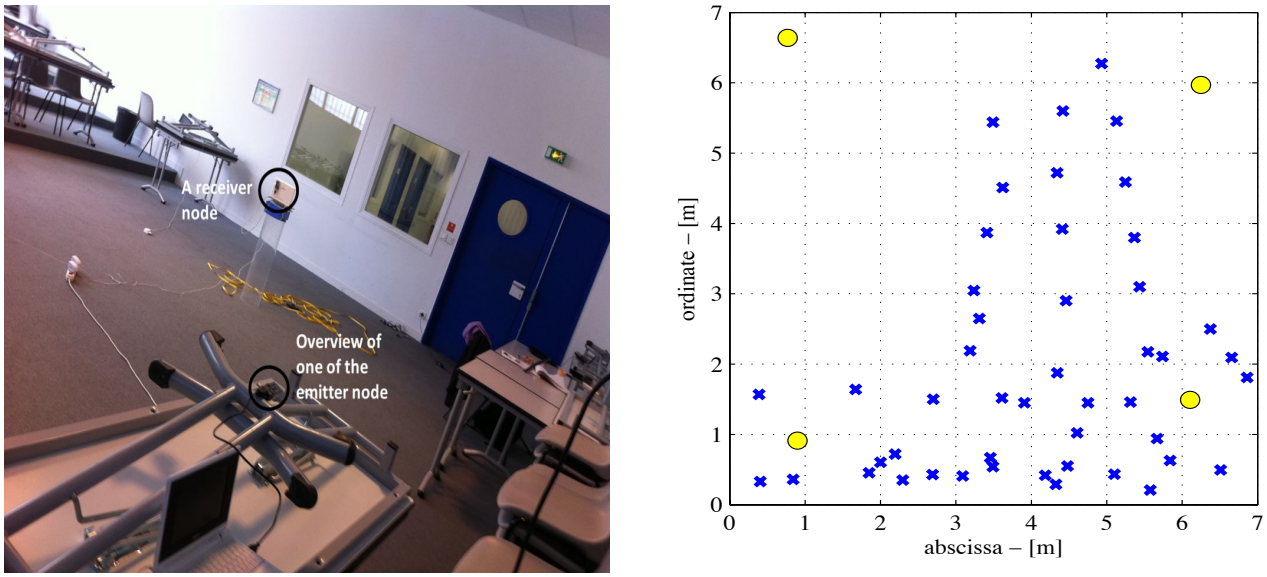
We present three principal testbeds in which experiments are carried out. Those testbeds are also designed to evaluate our proposed localization algorithms detailed in Chapter 4. Other locals and remote testbeds have been used for evaluating the characteristics of the indoor wireless link. Those are also presented below.

3.3.1 Details of testbed (#1) and (#2)

The first principal testbed named (#1) presented in Figure 3.2 is a semi-furnished office of size 4 m by 3 m. It is part of the Telecom ParisTech Lincs laboratory situated at place d'italie in Paris. It comprises 5 emitters. The office is divide into 6 squares grid and each square contains 8 random positions, for 48 receiver positions in total.

The second testbed named #2 is a semi-furnished classroom of size 7 m by 7 m. The testbed #2 is also part of Telecom ParisTech at Barrault site. Emitters in the number of 4 are placed over desired distances to cover the communication range over the 47 positions of the receiver. The receiver positions are randomly dispatched in 5 grids as shown in Figure 3.3. The number of sensor nodes in each testbed are chosen depending on the available hardware.

All emitter and receiver sensor nodes are Tmote Sky nodes and placed at the height of 1.25 m, which corresponds to the hip of the human.



(a) An overview of testbed (#2) of size 7 m by 7 m)

(b) Testbed 2: (4 landmarks, 47 receiver node unknown positions)

Figure 3.3: Locations of the landmarks (yellow bullets) and of the test positions (blue stars) in testbed (#2)

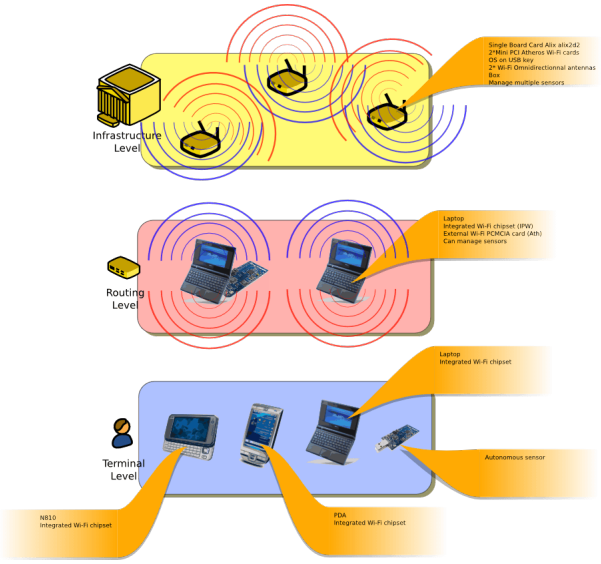
3.3.2 Details of testbed (#3)

The testbed (#3), illustrated in Figure 3.5 is a classroom with 8 emitters acting as landmarks. The number of landmarks in each testbed is chosen depending on the available hardware. In each environment, the receiving blind mobile node is positioned in different tested locations which are 57 in total. The testbed #3 measures 8.77 m by 6.46 m and is part of the Rennes Rammus platform hosted by the RSM department at Telecom Bretagne. Rammus stands for RAdio Mobile Multi Standards [139] and is a wireless experimentation platform deployed at Telecom Bretagne in both Rennes and Brest sites. Rammus enables the deployment in a real situation of algorithms using wireless networks for testing purpose. The Rammus platform is divided into three levels such as infrastructure, relay nodes and terminal nodes as shown in the diagram presented in Figure 3.4(a):

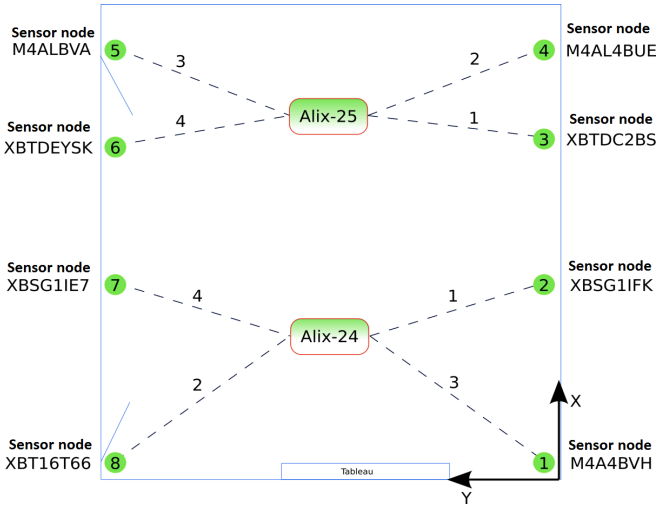
The infrastructure level is composed of ALIX [140] access points types that are programmable devices in GNU/Linux environment. Two Wi-Fi interfaces are available through a network of Ethernet and controlling at the same time one or several programmables 802.15.4 wireless sensor nodes [141]. For our experiments, we use the Rennes Rammus platform hosted by the RSM department of Telecom Bretagne, the one located in a classroom named *A27* and presented in Figure 3.4(b). This figure shows a top view of the testbed with 8 Tmote Sky nodes which can be programmable as emitters, receivers or both emitters and receivers functionalities.

3.3.3 Lincs Isup corridor testbed

The Lincs Isup corridor presented in Figure 3.6 is also located at Telecom ParisTech, Lincs laboratory in place d'italie, Paris. However it separates the Isup institute to the Lincs laboratory and is a busy corridor frequented by students. It is a hallway allowing access to the main meeting

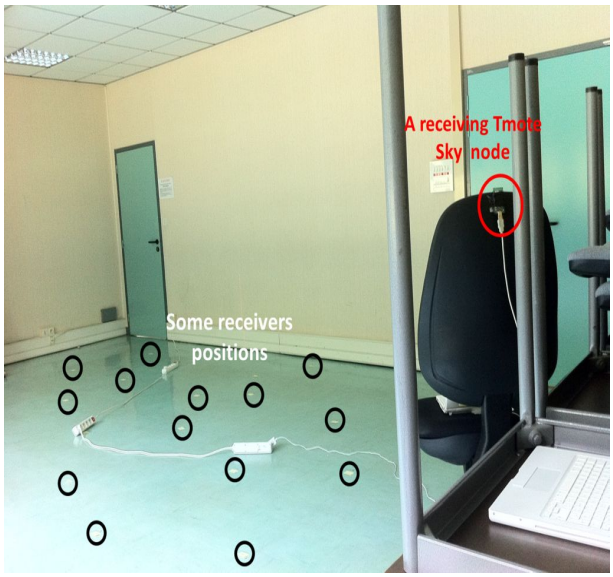


(a) The composition of the Rammus platform

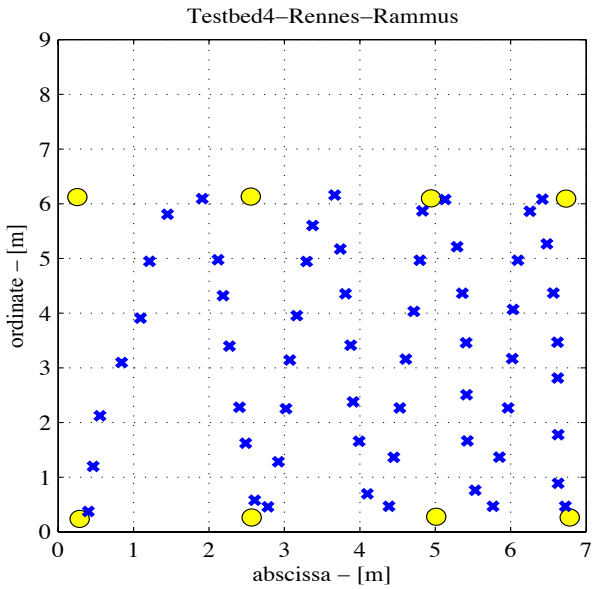


(b) Top view of the Rammus wireless sensor nodes platform at Rennes site

Figure 3.4: Rammus platform description



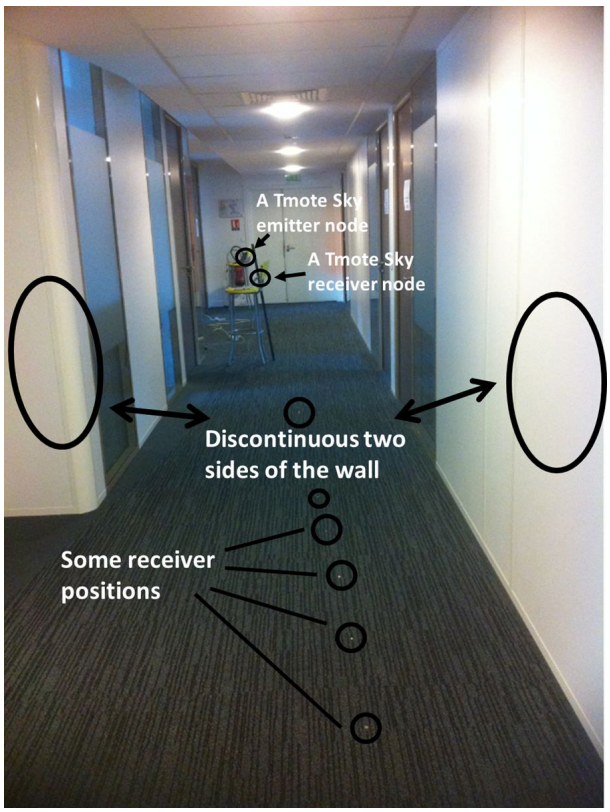
(a) An overview of testbed (#3) of size 8.77 m by 6.46 m)



(b) Testbed 3: (8 landmarks, 57 receiver node unknown positions)

Figure 3.5: Locations of the landmarks (yellow bullets) and of the test positions (blue stars) in testbed (#3)

(a) An overview of the Lincs Isup corridor



(b) Another view of the shape of the Lincs Isup corridor

Figure 3.6: Two views of the Lincs isup corridor

room. The dimension of the Lincs Isup corridor testbed is 2.54 m width, 2.30 m height and about 26.57 m long. During experiments, both emitters and receivers are placed at a height of 1.25 m from the ground.

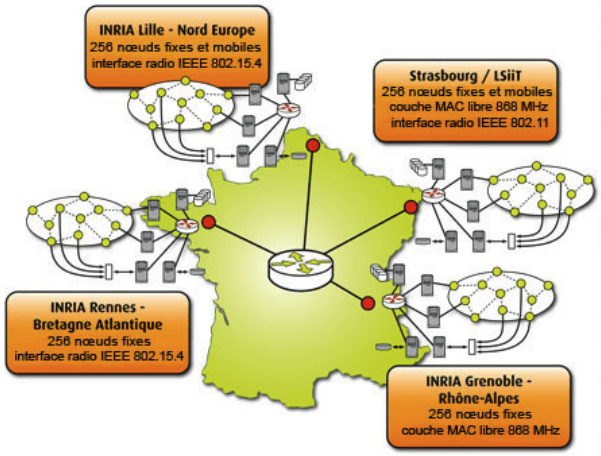
3.3.4 Description of the SensLAB platform

We additionally made some measurements in a remote large-scale wireless sensor network platform named SensLAB[2]. SensLAB is a wireless sensor platform for large-scale experiments involving communication protocols for sensor networks. SensLAB platforms are composed of 1024 wireless sensor nodes over 4 sites in France such as Grenoble, Lille, Rennes and Strasbourg [142, 143]. Figure 3.7 presents an overview of the SensLAB Rennes site which is essentially composed of 256 wireless sensor nodes using the CC2420 radio chip [82]. A frequency of 2.4 GHz radio interface is used as well as an omnidirectional SMD chip antenna [144] [144].

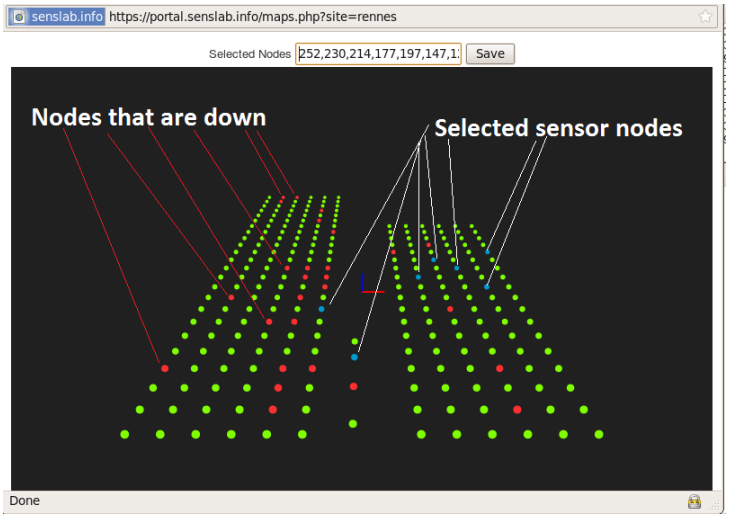
A SensLAB wireless sensor node named WSN430 is composed of three modules [144]:

- the open node running the user program.
- the control node storing data such as the level of the quality of the signal and the power consumption. However, we measure the RSSI values through a communication protocol that we have designed in *Contiki* operating system. Using the communication protocol, a

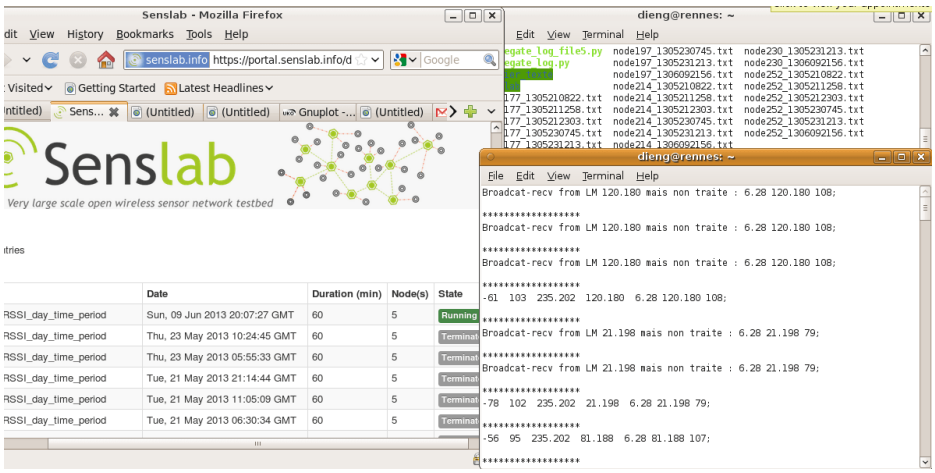
[2] `www.senslab.com`



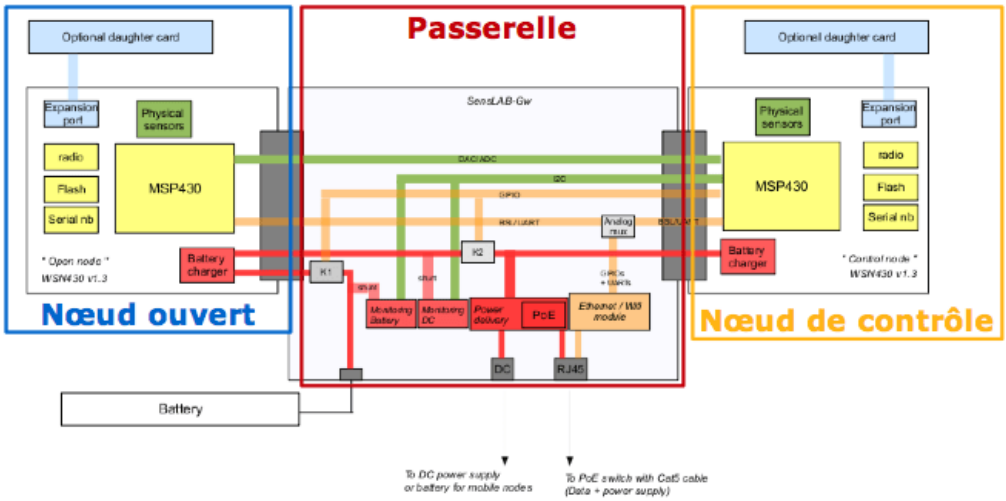
(a) SensLAB sites characteristics [145]



(b) Overview of the SensLAB Rennes WSN nodes platform during our experiment



(c) Overview of the application controller at SensLAB Rennes platform during our experiment



(d) An overview of SensLAB WSN430 wireless sensor node [144]

Figure 3.7: SensLAB plateform overview

Figure 3.8: An overview of the semi-furnished office testbed of size 7 m by 5 m (Testbed (#4))

receiver node is responsible for collecting needed information which includes RSSI values, emitters positions and sequence numbers of the received packets. At each location a receiver sends a localization request that is received by all emitters in its range. Each emitter then broadcasts beacons packets at a rate of one packet per second over a given duration and then stops emitting until another request is received.

- the gateway reprograming the nodes. The gateway embeds a small operating system and communicates with the server by its IP address. The gateway also controls the supply (i.e. battery) of the open node. A SensLAB WSN430 wireless sensor node is presented in Figure 3.7(d).

3.3.5 Other indoor local testbeds (#4) and (#5)

This paragraph describes two another testbeds. Testbed (#4), presented in Figure 3.8 is another office room of size 7 m by 5 m located at Telecom ParisTech Lincs laboratory. It is composed of 4 desks, cupboards, glass, and five people during working day time. The last testbed (#5) is another semi-furnished office room, located near the Lincs Isup corridor testbed. This last testbed has almost the same size and same configuration as the testbed (#4).

3.3.6 Experimental setup procedure

In the following, we attempt to analyze whether the RSSI with the use of the log-normal shadowing propagation model is a good indicator of distance between wireless sensor nodes in indoor environment In order to assure reliability of the obtained results, we used different indoor testbeds environments with different scenarios and a hallway testbed. A common setup procedure for each testbed has been explained above. Note that at least two persons are moving inside each testbed. To localize itself, a receiver node sends a localization request to emitters, which then begin sending broadcast packets at a rate of 256 kb/s. Such broadcast continues on the emitters node until about 100 packets are achieved and then stops. It does not reiterate until a next request comes from a receiver node. This process is repeated for all possible blind target nodes (i.e. receiver) positions. Additionally, the 2.4 Ghz frequency is used, which is the

same used in 802.11 wireless network. During experiments, we remark the presence of such wireless networks. The type of sensor nodes we used are Tmote Sky nodes from Moteiv Corporation [131]. Sensors are programmed using Contiki 2.5 operating system release[3]. Each node is powered through its USB port to an *Asus Eee pc* computer, which eliminates any decrease or variation coming from battery power levels. Furthermore a manual rangefinder was used for all physical measures; this can imply a small difference in terms of precision. This part of our work is intended for studying the reliability of the measured RSSI values from the sensed data packets. In fact, our aim is to accurately position a blind mobile user/object that carries a wireless sensor node.

3.4 Characterization of the Tmote Sky mote and of the RSSI metric

As we mention, the received signal strength indicator has the advantage of requiring no additional hardware. It is already implemented in Tmote Sky sensor nodes [131] and is widely used in 802.11 and in 802.15.4 standards. Viewed as an indication of the incoming signal power level being received by the antenna, each RSSI value is a voltage measurement according to a common frequency [146]. Such frequency is provided for each received radio message by the CC2420 radio chip module supplied with a sensor node [82]. Note that RSSI readings via CC2420 have a dynamic range of about 100 dBm.

In embedded devices like in TMote Sky nodes from Moteiv Corporation, the received signal strength is converted to RSSI and is defined as ratio of the received power to the reference power of 1 milliWat as:

$$\text{RSSI}(dBm) = 10\log_{10}(\frac{P_r}{P_{ref}}) \tag{3.1}$$

Where P_r is the signal reception power and $P_{ref} = 1mW$. The RSSI computation built in CC2420 produces a discrete value averaged over 8 symbol periods and stored in a register named RSSI VAL. Chipcon specifies the following formula to compute the received signal power (i.e. P) in dBm: $P = RSSIVAL + RSSIOFFSET$, where $RSSIOFFSET \approx -45$. We refer to this power, named P (in dBm) as RSSI throughout this document.

3.4.1 Characteristics of a Tmote sky mote in an anechoic chamber: Tmote Sky antenna radiation pattern

This section analyses the antenna effects on the RSSI fluctuation. In the following we draw the antenna diagram of a TMote Sky node in an anechoic chamber to check whether the antenna gain is, as announced in the datasheet, close to almost omnidirectional but not perfect, i.e. the gain is somehow independent to the direction. A front back view of the Tmote Sky node is presented in Figure 3.9.

To determine the polar variation of the emitter w.r.t the receiver deployment, we first test the antenna polarization by having the emitter turning around an horizontal axis. We measure the partial gain of the Tmote Sky node with respect to a vertical and an horizontal deployment

[3] `https://www.sics.se/search/content/contiki`

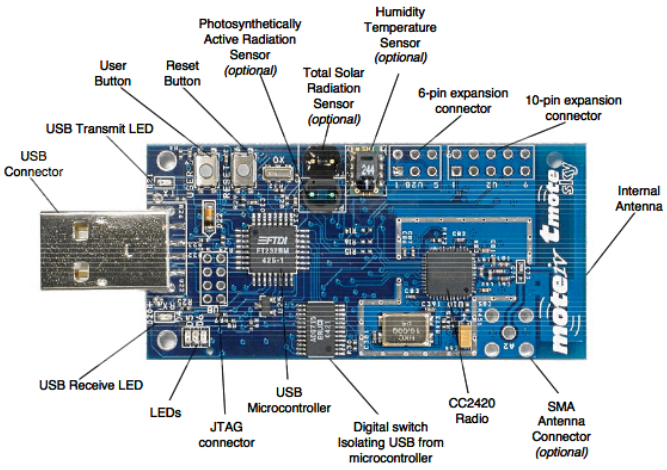
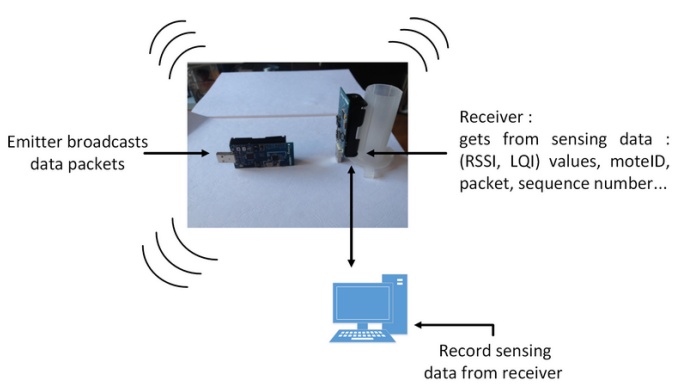
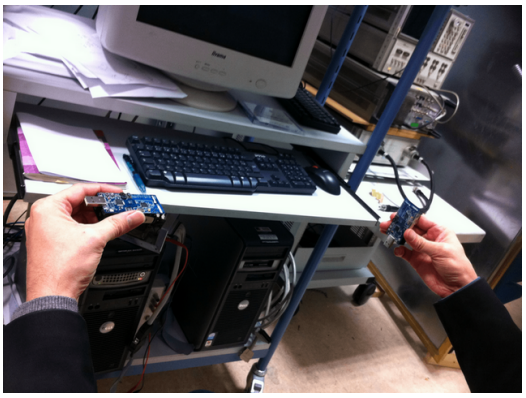


Figure 3.9: Block diagram and front view of the Tmote Sky node module [131]



(a) Vertical deployment of a receiver



(b) Vertical deployment of a receiver: another view

Figure 3.10: Deployment position scenario A : a vertical deployment of the receiver w.r.t an horizontal deployment of an emitter rotated within horizontal axis

position of the receiver node and when the emitter is deployed horizontally. We first study a vertical deployment of a receiver represented by the deployment position scenario named A (cf. Figure 3.10). We measure RSSI values at the receiver side when the emitter is only rotated around the horizontal axis from 0° to 350° with a step of 10° as shown in Figure 3.11. The results in Figure 3.11 show that this polarization is uniform. In fact, Figure 3.11 reports less variation in polarization (i.e. polar without variation).

The variation of the emitter within the horizontal axis is also performed with respect to the horizontal deployment of the receiver as shown in Figures 3.13 and 3.12. Experiment results show some variations in polarization (i.e. polar with the presence of variations) when the receiver lies horizontally. To resume, a better polarization and less variation is reported within the vertical deployment of the receiver w.r.t the horizontal deployment of the emitter node (represented by the deployment position scenario A).

To confirm both hypothesis : the not omnidirectional pattern of the Tmote Sky node and the not adapted horizontal deployment position of the receiver with respect to the horizontal deployment of the emitter, we measure the total gain of the Tmote Sky antenna for an horizontal deployment

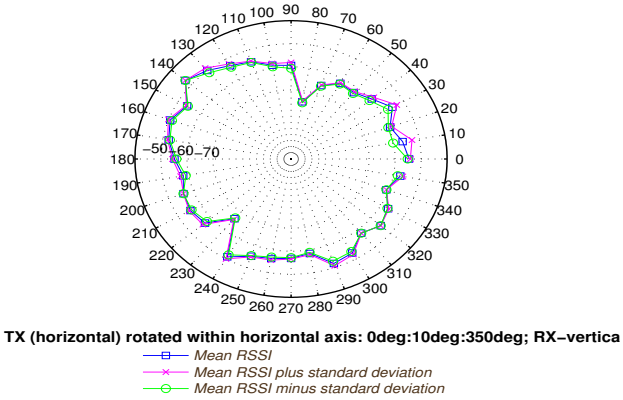


Figure 3.11: Emitter i.e. TX (horizontal) rotated within horizontal axis w.r.t a vertical deployment of a receiver

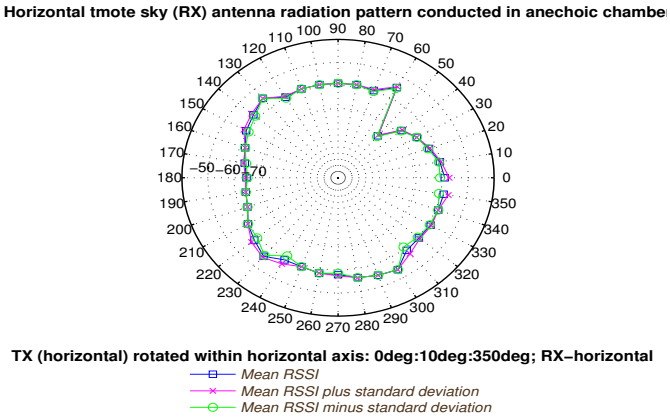


Figure 3.12: Emitter i.e. TX (horizontal) rotated within horizontal axis w.r.t horizontal deployment of a receiver

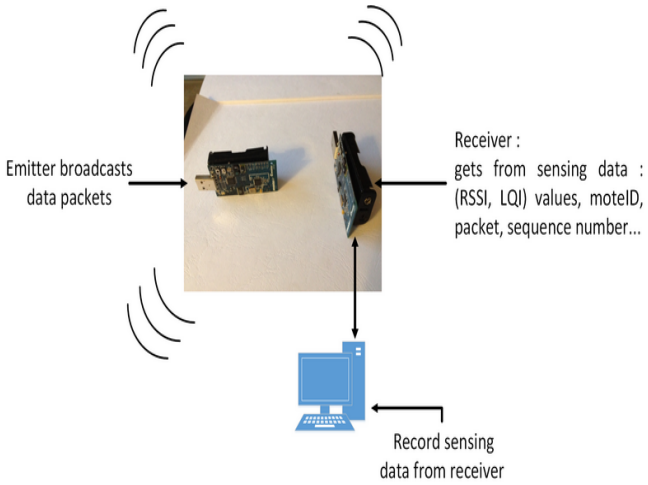
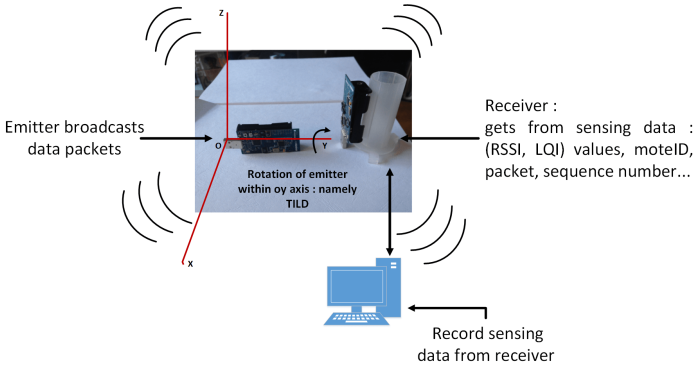
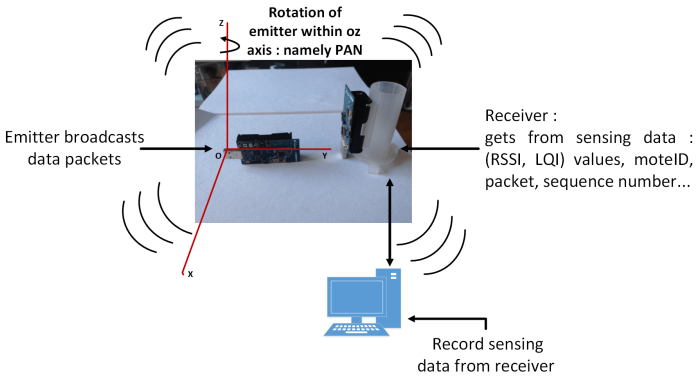


Figure 3.13: Deployment position scenario B : an horizontal deployment of a receiver w.r.t a horizontal deployment of an emitter rotated within horizontal axis



(a) Rotation of an emitting antenna within horizontal axis



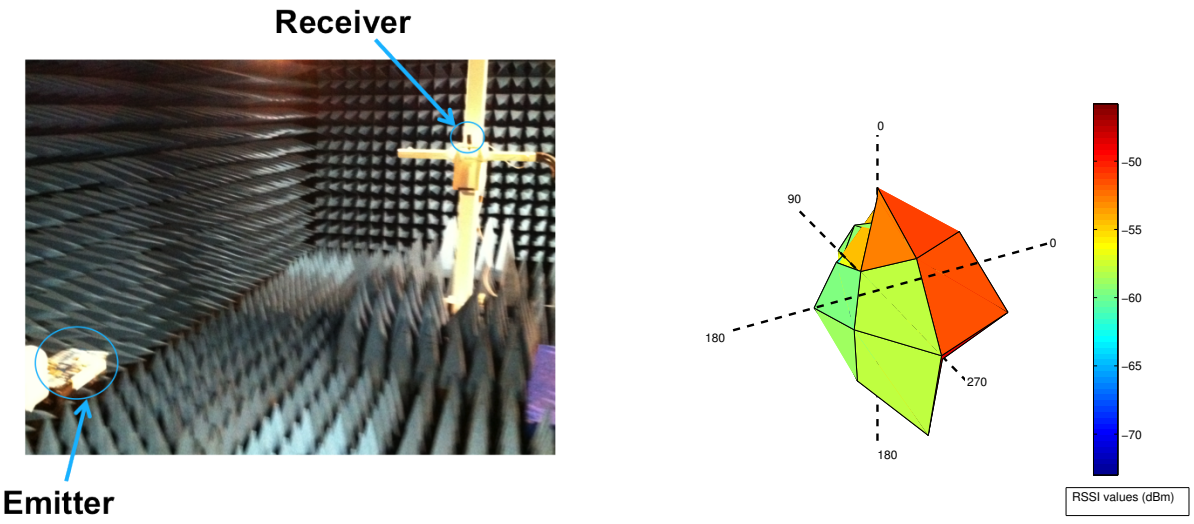
(b) Rotation of an emitting antenna within vertical axis

Figure 3.14: Architecture of a 3D radiation pattern of a Tmote Sky node conducted in an anechoic chamber

of a receiver with respect to an horizontal deployment position of an emitter antenna. As shown in Figure 3.14 the emitter antenna is first rotated within horizontal axis w.r.t these angles : 0°, 45°, 90°, 135° and 180°. Second, in each of these angles the same emitter antenna is rotated with respect to the vertical axis ranging from 0° to 360° with a step of 45°. Finally, for each deployment position of the emitter, we have collected the RSSI values on the receiver being deployed vertically (cf. Figure 3.14). The distance between the emitter and the receiver node is set to 2.30 m and there are no obstacles that introduce reflections, fading within this separation distance. The emitted power is set to 0 dBm and all data are measured during 10 seconds for each angle. The $3D$ radiation pattern is then plotted and represented in Figure 3.15.

Experiments results of the Tmote Sky radiation conducted in an anechoic chamber at Telecom ParisTech do not show a directive diagram, but also not an omnidirectional pattern. An omnidirectional antenna is an antenna which radiates in same way in all directions or with very slight variations; this is not the case on the drawn diagram for Tmote Sky node. Such results confirm those obtained in the works of Sungwon Yang *et* al. [147] in which it is assessed that the common $\frac{1}{4}$ wavelength monopole antenna which is supplied with a Tmote Sky node does not have a regular pattern.

In addition, the authors have demonstrated that antenna orientation problem comes from the small length of the ground plane. A second factor affecting the radiation pattern of a PCB internal antenna is the closeness to circuits as well as to devices that compose a sensor PCB. In fact, the electric field of the antenna is disturbed by interference from other devices like



(a) An overview of the Telecom ParisTech anechoic chamber

(b) 3D radiation pattern of the Tmote Sky conducted in the Telecom ParisTech anechoic chamber

Figure 3.15: 3D radiation pattern of the Tmote Sky node conducted in an anechoic chamber at Telecom ParisTech

SubMiniature version A and peripheral connectors that are close to the antenna. This can be solved by matching the ground plane with the proper length of the antenna or even keeping enough distance between an antenna and sensor node peripheral devices to minimize the electrical interference from devices of a sensor node and keep the horizontal omnidirectional radiation pattern of the internal antenna. Furthermore, the authors (i.e. in [129]) experimental results using 1.5 m separation distance between two wireless nodes and an output power of −10 dBm have demonstrated that a monopole antenna with a wavelength $\lambda = \frac{1}{2}$ or $\lambda = \frac{5}{8}$ has a more isotropic radiation pattern than a common wavelength $\lambda = \frac{1}{4}$. The $\frac{1}{2}$ wavelength monopole horizontal radiation pattern is omnidirectional. These results are also affirmed in an outdoor case by the works of Graefenstein *et al.* [148]. This radiation pattern becomes more perfect circular for both the $\frac{1}{2}$, $\frac{5}{8}$ and $\frac{1}{4}$ wavelength monopole when the antenna is mounted a wavelength away from a sensor mote. In case of mounting, the radiation pattern is more circular with the $\frac{1}{2}$, then follows the $\frac{1}{8}$ and then the $\frac{1}{4}$ wavelength monopole antenna; however, RSSI values get weaker when compared to the $\frac{1}{4}$ wavelength. From the authors, this is due to the fact that, the $\frac{1}{2}$ or $\frac{5}{8}$ wavelength monopole length do not match the given frequency of 2.4 GHz i.e. antenna does not resonate at the given frequency of 2.4 GHz. Therefore, such wavelength receives weak signals, and consequently shortens the communication range. Thus, an antenna that has an analogous radiation pattern to $\frac{1}{2}$ or $\frac{5}{8}$ wavelength antenna without attenuation of RSS is required. However, in this thesis, the question to solve is how to overcome and deal with such irregularities when it comes using the manufactured Inverted-F $\frac{1}{4}$ wavelength monopole internal antenna.

Ideally, it would be better to have a range of values where the directivity of the $\frac{1}{4}$ wavelength monopole antenna is fairly constant for a largest angular range as possible, hence an omnidirectional radiation pattern. However, our following experiment results present lots of dispersions on the measured RSSI values. RSSI measurements conducted in an anechoic chamber show the right part in Figure 3.15 to present a better polarization when the rotated angle around the ho-rizontal axis is around 0 °, 90 ° and 270 ° for any emitter with respect to the deployment position as in left side of Figure 3.15. Over −50 dBm is observed there, which is more better than around 180 ° in which we observe RSSI readings between −60 dBm to −55 dBm. Furthermore,

when looking angles that are rotated through the vertical axis, we notice RSSI values being around −50 dBm and between −60 dBm to −55 dBm at 180°. We can note that the radiation pattern is not entirely uniform, nor smooth, a few directions experience a strong attenuation due to electronics on one of the board as shown in Figures 3.15 and 3.16.

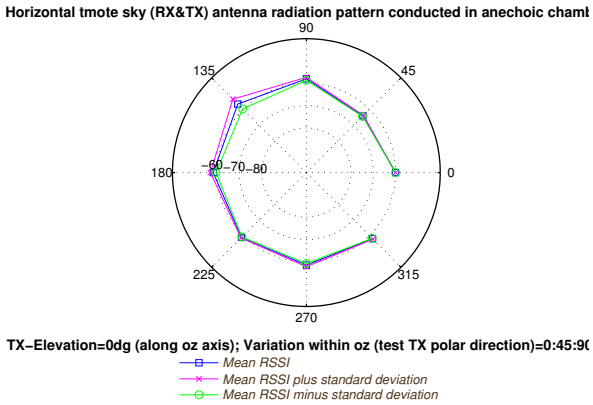
We come up with the conclusion that in short, what is important is the relative orientation of the antennas one with regards to the other. In addition, a more uniform polarization and thus less variation is observed for a vertical deployment of a Tmote Sky receiver w.r.t an horizontal deployment of a Tmote Sky emitter node (cf. Figure 3.11) compared to an horizontal deployment of the same receiver. The horizontal deployment of a Tmote Sky receiver (cf. Figure 3.12) reports some variations in polarization. In brief, Information from datasheets should in most cases be verifiable. Electronics devices we manipulate are not almost, nor fully omnidirectional. Such results confirm our guess and would be considered for all kind of devices. Our algorithms we design fully consider such results. In our subsequent experiments, we take into account such deployment scenarios of sensor nodes in order to better interpret ensuing results. In the next paragraph, we consolidate these results by measuring RSSI values using the same node deployment scenarios in a typical indoor scenario.

3.4.2 Characteristics of a Tmote sky mote in a typical indoor environment

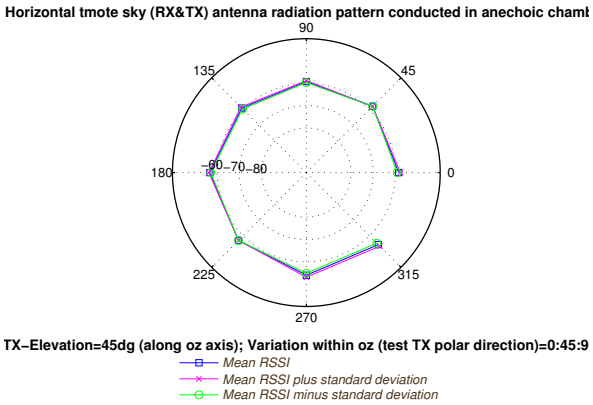
The analysis of the signal strength variability by means of the relative position of the emitter-receiver antennas deployment is necessary in indoor wireless environment due to its complexity that renders indoor location more difficult. To this purpose and to improve ranging accuracy, thorough measurements are needed in several indoor environments with different configurations to study the behavior of the RSSI. We do not use an isolated room to be resilient to interference/perturbation. Our approach is different to the experiment condition performed in the works of Lymberopoulos *et al.* [133]. In fact in such works, the wireless nodes under test were placed in the middle of a room without furniture to minimize the effects of reflections in measurements.

We first perform measurements between 4 emitters and one receiver using different positions of deployment of emitters-receiver with the default integrated PCB antenna. These inclinations are different from those presented in the works of Stoyanova *et al.* [149] where nodes deployment were only based on emitters antenna orientations. The receiver antenna should also be taken into account in terms of which emitter position provides less variation on reported RSSI values in accordance with a receiver antenna. As presented in Figures 3.17 and 3.18, in all of our experiments, the receiver are exactly in two positions (i.e horizontal and vertical deployment) while the same antenna orientation is used for each of them. In each receiver position of deployment (cf. Figures 3.17 and 3.18), we study two positions of deployment of the emitters which are in the total number of 4. The four considered deployment positions with respect to both the emitter and receiver antennas are respectively named *TowardCenter emitter parallel receiver*, *FromCenter emitter parallel receiver*, *TowardCenter emitter perpendicular receiver* and *FromCenter emitter perpendicular receiver* as shown in graphs of Figures 3.17 and 3.18. For simplicity, we will call them respectively: *TowardCenter parallel*, *FromCenter parallel*, *TowardCenter perpendicular* and *FromCenter perpendicular*. The *parallel* one considers both, the emitters and a receiver antenna to be parallel and in horizontal position.

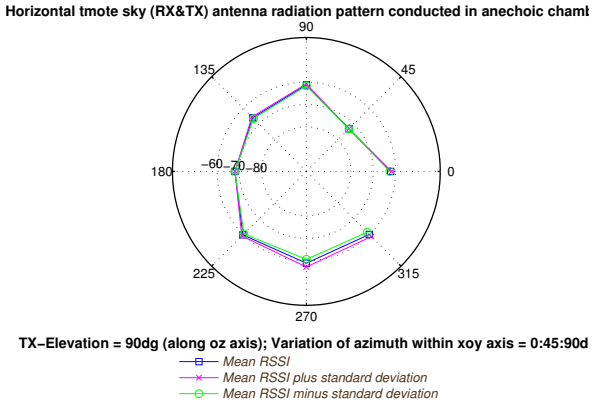
More precisely, considering the graphs in Figures 3.17 and 3.18 the term *TowardCenter* lies to the fact that the emitter antenna is in horizontal position and pointed inside the room, the sensor board is also in horizontal position. The battery pack facing the wall and the USB pointing



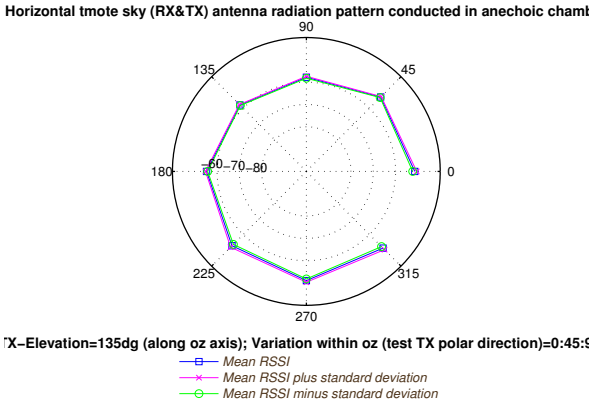
(a) Emitter rotated in the horizontal axis: 0° and in the vertical axis from 0°:45°:360°



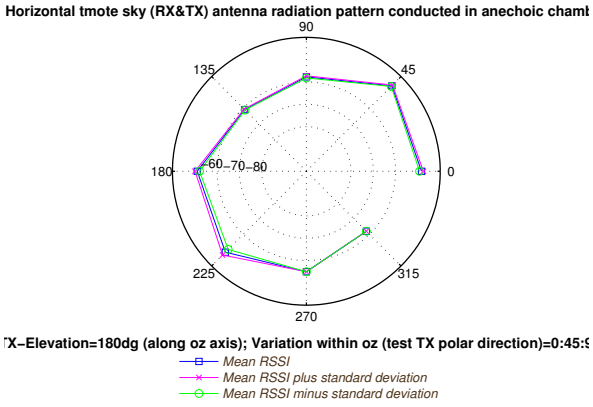
(b) Emitter rotated in the horizontal axis: 45° and in the vertical axis from 0°:45°:360°



(c) Emitter rotated in the horizontal axis: 90° and in the vertical axis from 0°:45°:360°



(d) Emitter rotated in the horizontal axis: 135° and in the vertical axis from 0°:45°:360°



(e) Emitter rotated in the horizontal axis: 180° and in the vertical axis from 0°:45°:360°

Figure 3.16: 2D antenna diagram in an anechoic chamber representing RSSI values vs. antenna orientation when emitter is rotated considering horizontal deployment with respect to an horizontal deployment of the receiver

toward outside of the room.

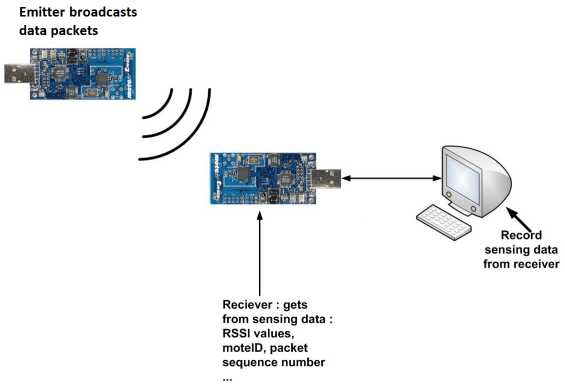
The graphs in Figures 3.17 and 3.18 show the deployment scenario of the antennas as well as the RSSI values recorded at the receiver at each of the two receiver deployment inclinations (i.e horizontal and vertical deployment) with respect to both emitter deployment inclinations. Each emitter was broadcasting packets in two different orientations.

The term *FromCenter* refers to an emitter antenna pointing outside the room, its battery pack facing the wall and the USB part points inside the room. Instead, the term *perpendicular* corresponds to a vertical position of a receiver: the USB pointing to the ground and the antenna up towards the ceiling as shown in graphs in Figures 3.17 and 3.18.

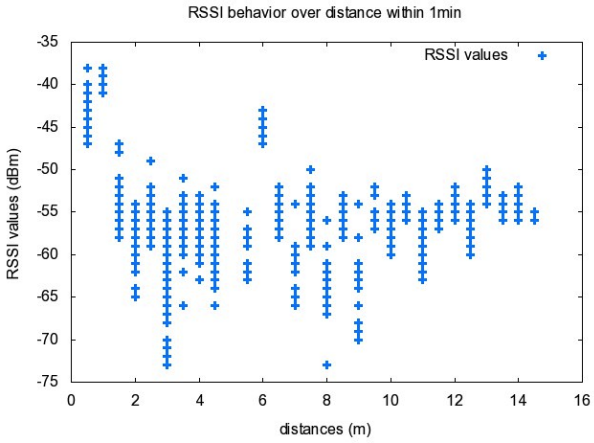
In each position of the receiver, we took about 140 measurements which are about one minute interval; packets are broadcasted randomly by emitters at each second. RSSI values reported on a receiver are then plotted according to an emitter. The corresponding results present different curves shapes according to each architecture; but they still all have the same trends. We always see that the RSSI values are not well correlated with distance. However, the architecture for *TowardCenter perpendicular* antennas seems to be better as it shows less dispersion of RSSI values in the LINCS-ISUP corridor testbed. Results from the architecture of the *TowardCenter perpendicular* antennas also present a more Gaussian pattern compared to both results from the *TowardCenter parallel* antennas and the *FromCenter perpendicular* antennas in which the obtained curves resemble to two non overlapping Gaussian bell curves.

In addition, the different positions of deployment of the sensor nodes do not produce the same polarization results. A better polarization is observed with the *TowardCenter perpendicular* antennas deployment (from -60 dBm to -30 dBm) compared to the other three deployment positions which report less polarization such as $[-75dBm - 35dBm]$ for the *TowardCenter parallel* antennas, $[-75dBm - 30dBm]$ for the *FromCenter parallel* antennas and $[-65dBm - 25dBm]$ for the *FromCenter perpendicular* antennas. Indeed, the *TowardCenter perpendicular* antennas deployment shows less fluctuations (i.e less difference in dB between the RSSI values) compared to the others. Our actual results achieved by several internal environments are consistent with those of Azevedo *et al.* [150] realized inside a laboratory with non-mentioned size using a Tmote Sky node as an emitter and a spectrum analyzer acting as a receiver in a line of sight environment. Works in [150] also demonstrate the presence of variations between RSSI readings at the same position when using different transmitted power. The authors also show that two types of sensor nodes can not produce the same polarization results as a difference of about 13 dBm was reported on the signals measured from both: Tmote Sky and Micaz sensor nodes. Furthermore, it is demonstrated that due to ground reflection, the distance of the sensor node from the ground should be greater than 40 cm; under this height, RSSI began fluctuating. Almost the same variation of the received signal according to distance was observed for receiving antennas placed at 40 cm and 1 m above the floor.

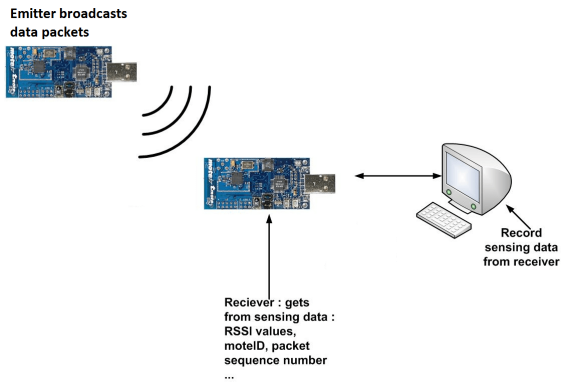
The curves obtained by this first experiment outside anechoic chamber do not only confirm the shape of the radiation pattern of the Tmote Sky sensor node, but also identify one cause of interference due to deployment positions of an emitting node with respect to a receiver node. Further experiments in different indoor environments are going to be performed in the following to better characterize the variations of the RSSI. For this reason the measurement results about the indoor obstruction cause proposed in [133, 147, 134] were adopted while adding other effects presented in Section 3.4.3 and a more dynamic procedure in the way such effects can be mitigated during measurement process. Results from our experiments and conclusion from these works [133, 134, 135] conduct to adopt these following parameters in our work:



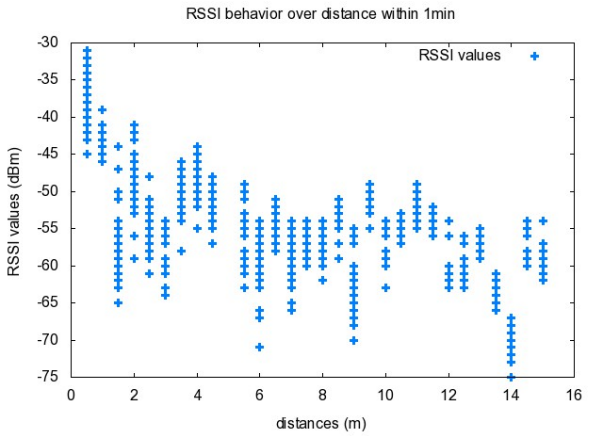
(a) TowardCenter parallel antennas scenario (i.e A)



(b) Reported RSSI values from the scenario A

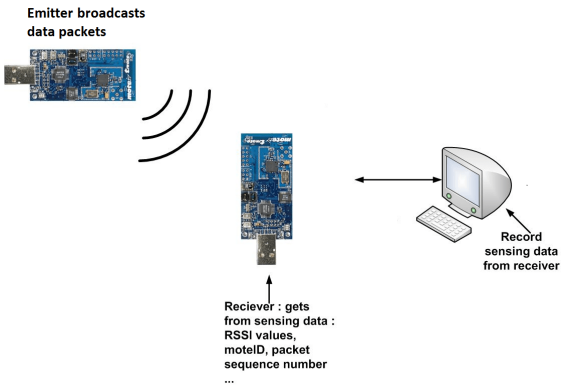


(c) FromCenter parallel antennas scenario (i.e B)

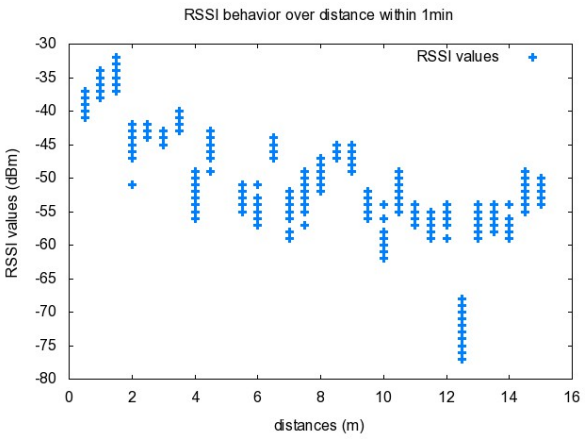


(d) Reported RSSI values from the scenario B

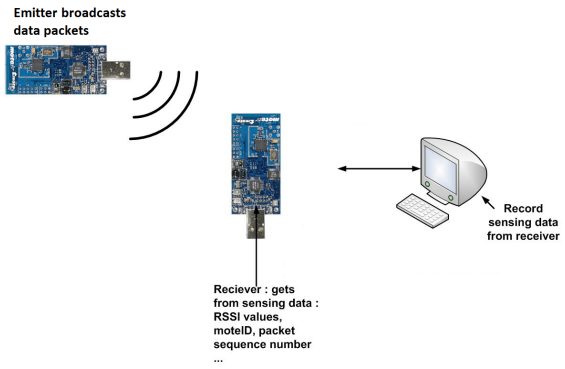
Figure 3.17: RSSI distribution from the Lincs-Isup corridor performed at different emitter-receivers deployment positions (part 1)



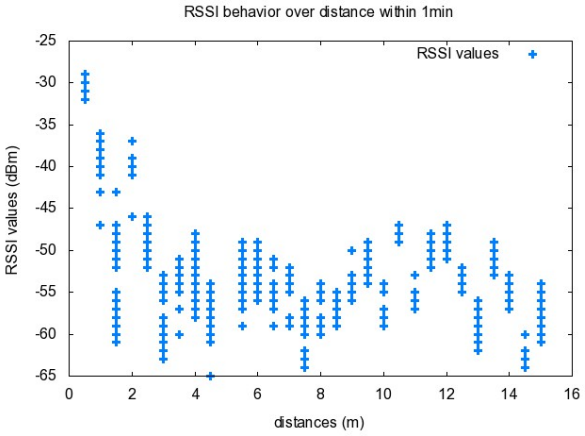
(a) TowardCenter perpendicular antennas scenario (i.e C)



(b) Reported RSSI values from scenario C



(c) FromCenter perpendicular antennas scenario (i.e D)



(d) Reported RSSI values from scenario D

Figure 3.18: RSSI distribution from the Lincs-Isup corridor performed at different emitter-receivers deployment positions (part 2)

- Exploiting channel utilization strategy from the results observed in [135] in which the authors come up with the experimental conclusion that the channel 26 has the best quality and reduces packet loss rate. Therefore the channel 26 is used for node communication. Besides, the average RSSI value increases as the transmission power enhances. We also know that the 802.15.4 channel 26 does not overlap with any 802.11b/g channel [35], which minimizes interference from co-located Wifi networks.

- The output power 0 dBm is used to cover the whole experiment area and to follow the previous efforts from [123]: a lower transmission power does not conduct to a good range in distance measurement. In fact, levels of transmission power lower than 0 dBm or power level 31 [99] render a problematic communication because of the great percentage of packet losses: more than 80% [134].

- The following node alignments are used: a vertical position of deployment of a receiver and an horizontal position of deployment of emitters through the deployment scenario C called the *TowardCenter perpendicular* antennas.

- The emitters and receivers antennas are placed at the same height as recommended by Lymberopoulos *et al.* in [133] and Yang *et al.* [129]. In addition, Stoyanova *et al.* [134] have demonstrated that higher heights from the ground conduct the RF signal to be more sensitive to the reflections than the lower heights which present more "energy holes". In this way, the height from the ground should not be more than 2 m in order to correspond to the standard height of the human and being minimum to 1 m [133]. Under the given condition, the height of 1.25 m from the ground is used for all Tmote sky nodes involved in our experimental work. Furthermore, the increase of the height difference between an emitter and a receiver antennas leads the antenna orientation to become a principal cause of the signal propagation perturbation [133].

- Two persons are moving inside each environment and impact radio propagation. A few Wi-Fi access points are also located nearby in order to stand in real experimental conditions like a room, a confined space in indoor building.

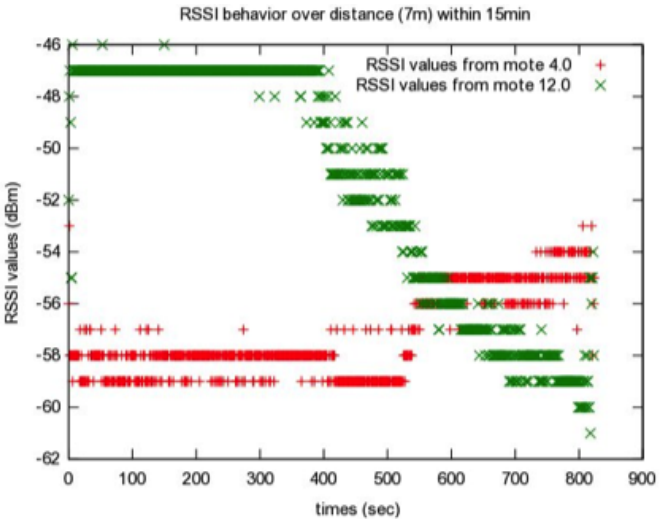
We should also notice in the works of Lymberopoulos *et al.* [133] and of Stoyanova *et al.* [134] that some frequencies are more vulnerable to disturbance than others. In fact a higher frequency is better as a high standard deviation was remarked on the lower frequencies.

In the next paragraph, we will study the other reasons of the RSSI fluctuations by investigating the relations among RSSI and distance. Moreover, it is interesting to investigate the path loss effect at different environments with interference and obstacles.

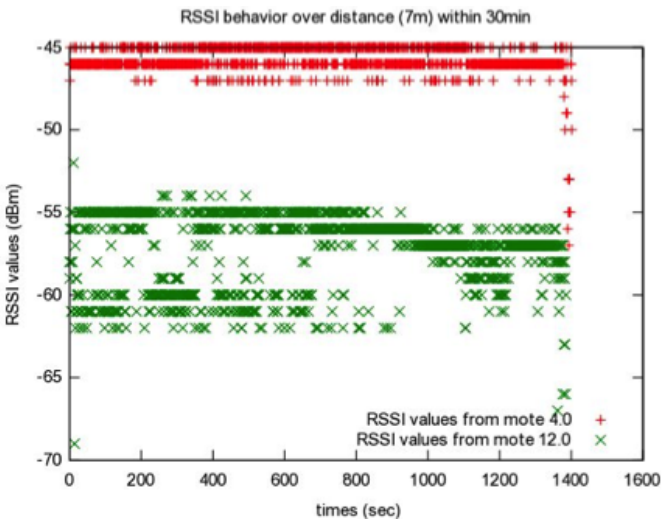
3.4.3 RSSI to distance mapping

The linearity of RSSI readings with respect to a corresponding distance To test whether the RSSI values are linear for the same distance, we performed some measurements with two equidistant emitters and one receiver in a corridor whose details are given in Section 3.3.3. Data coming either from emitter #1 (node $identifier = 4.0$) or emitter #2 (node $identifier = 12.0$) are collected at the receiver in two different intervals of 15 min and 30 min respectively. Results are shown in Figure 3.19.

Figure 3.19 shows the RSSI values recorded at one receiver for both emitters within the same orientation. In such figure, the first emitter node $identifier = 4.0$ is represented by "Plus sign"



(a) Experiment conducted between one receiver and two equidistant emitters during 15 min

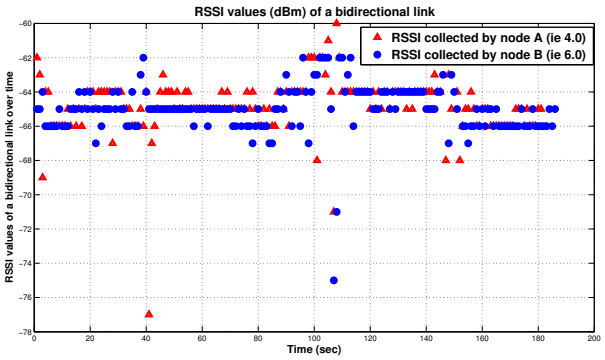


(b) Experiment conducted between one receiver and two equidistant emitters during 30 min

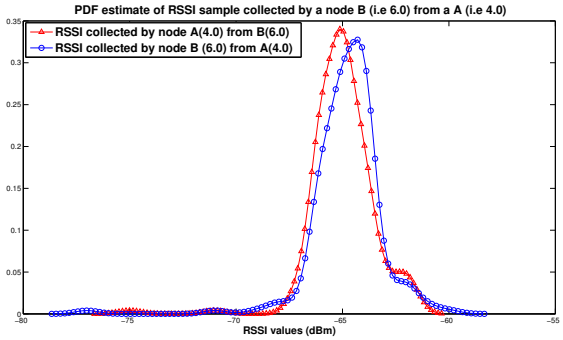
Figure 3.19: Experimental results when testing the RSSI to distance reliability from the LINCS ISUP corridor: first emitter node $identifier = 4.0$ represented by plus sign (+) and second emitter node $identifier = 12.0$ represented by crosses (x)

(+) and the second emitter node $identifier = 12.0$ by crosses (x). The used orientation is the one described in Figure 3.18 which is the *TowardCenter perpendicular* antennas deployment scenario. Analyzing all the recorded RSSI values over time for both emitters, we found that the distribution of the RSSI values from the equidistant emitters are different according to the receiver. The results highlight the presence of different asymmetric links. In addition, the environmental location of the emitter is important and the indoor channel varies with the time. In fact, the two emitters located at equal distance of 7 m to a receiver do not result in the same RSSI behavior. We realized the same experiment several times with different emitters and receiver. For each experiment we did not measure the same RSSI behavior regarding each equidistant emitter with respect to a receiver. Such results identify a variability on RSSI readings depending on the emitter environmental location. Besides the dispersion of RSSI values is not equal in two different time periods. These variations can be attributed to obstacles arising from people moving inside the environment, attenuation from the ceiling, floor as well as the shape of the wall constituting the corridor.

Is the communication of wireless sensor node symmetric or asymmetric? In this paragraph, we study the properties of bidirectional communication links between sensor nodes. Indeed a knowledge of the links properties (i.e symmetric or asymmetric) is necessary as it is an information that contributes during the design of a localization algorithm. In the case of a communication scenario involving sensor nodes with both functionalities (emission and reception), we draw for each individual link the distribution of the RSSI readings. We additionally draw the density function of the measured RSSI values from the data packets sensed by each wireless sensor node receiver in testbed (#1) (cf. Section 3.3.1). We also compare the shape of all density functions to a normal distribution. Such comparison is made using the Matlab normal probability density function applied to our measured RSSI samples. Compared density functions are computed from the measured data to be analyzed and results are represented in Figure 3.20 in which we show almost symmetrical radio links between the two nodes.



(a) Distribution of the RSSI values from each individual link



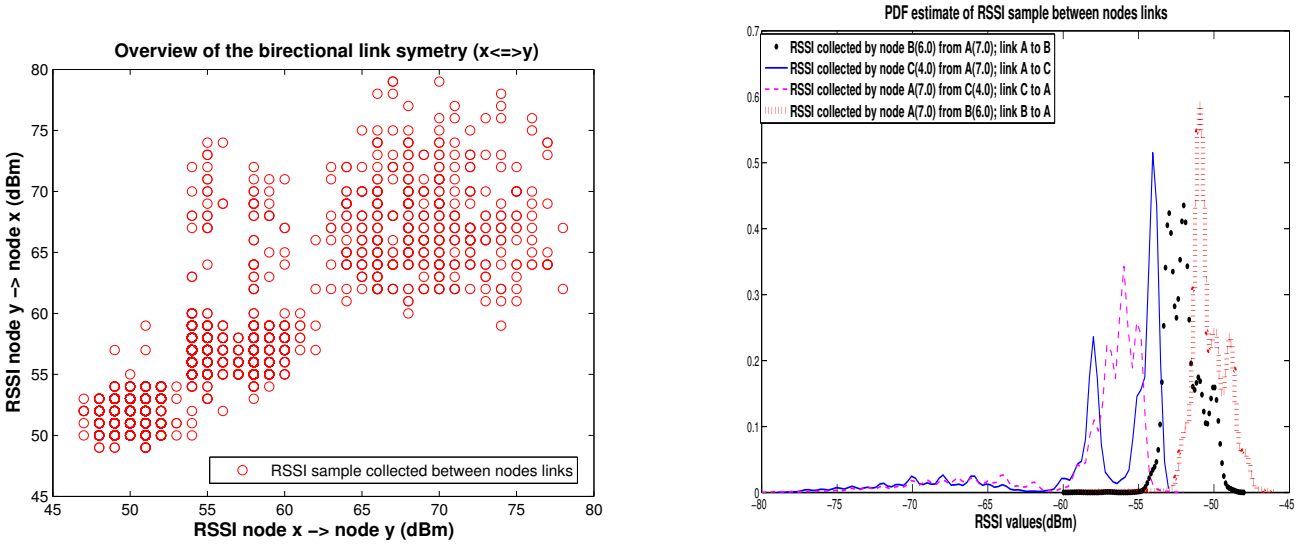
(b) Density function (PDF) estimated from RSSI sample of a bidirectional link

Figure 3.20: Overview of a symmetrical bidirectional link between two wireless sensor nodes

However, we do not observe the same results when we further consider another two-unidirectional link involving two wireless sensor nodes labeled A (for node identifier 7.0) and C (for node identifier 4.0) as shown on the right curve in Figure 3.21. From communication links between A (node identifier 7.0) to C (node identifier 4.0) and C (node identifier 4.0) to A (node identifier 7.0), the RSSI readings are generally almost centered at the same average value of -56 dBm. However, we report a totally different representation accompanied by lots of disruption between both links curves.

Indeed, the curve with solid line in blue color in Figure 3.21 and representing the unidirectional link A (node identifier 7.0) to C (node identifier 4.0) has two independent maximum values on -58 dBm and -54 dBm. This is not the case for the curve with dashed line in pink color for unidirectional link C (node identifier 7.0) to A (node identifier 4.0). This latter shows small and close peaks of different sizes: a maximum value at -56 dBm with two small peaks at -55 dBm and -57 dBm. In the curve of the communication link A (node identifier 7.0) to C (node identifier 4.0) (i.e. curve with solid blue line) we note two distinct maxima and a large disturbance of about 4 to 5 dBm difference between them. Thus, the signal has been received two to three times larger than the emitted signal between these two maxima on that unidirectional link. However, for the communication link of C (node identifier 7.0) to A (node identifier 4.0) represented by the curve with dashed pink line, there is a difference between its maxima called delta signal equal to approximately 1.2 times higher than the output power. Such difference can be somehow negligible.

From the overall relationship between the two communications links: A (node identifier 7.0) to C (node identifier 4.0) and C (node identifier 7.0) to A (node identifier 4.0), when taking the average value of the maximum for each link, we are left with -56 dBm on both sides and the resulting difference leads to a signal received with the same output power. The presence of significant disturbances on the signals is also noticed. Such effects result in asymmetric links (i.e links A (node identifier 7.0) and C (node identifier 4.0)). Unidirectional links C (node identifier 4.0) to A (node identifier 7.0) and A (node identifier 7.0) to C (node identifier 4.0) are roughly centered at the same averaged mean values, but distributions are too different and there is too much interference between the two curves. Considering the curve with solid blue line, we observe the presence of two independent peaks or two maxima, this is not the case for the curve with dashed pink line. This curve with dashed pink line presents different peaks of small size and enough close. There is also an asymmetry behavior for the link A to B (node identifier 6.0) with



(a) Distribution of the RSSI values from several individual links (b) Density function (PDF) estimated from RSSI sample of several bidirectional links

Figure 3.21: View of symmetrical bidirectional links

the dotted line curve and for the link B to A (node identifier 7.0) with dash-dotted line. Such links have small peaks and a difference of 2 to 3 dBm between their maxima.

From one link to another we can observe a symmetrical or asymmetrical behavior depending on disturbances on the signal and at the same time on the measured RSSI values. Therefore we can not anticipate the behavior of the intensity of the signals received at two different unidirectional radio communication links as disturbances can provoke asymmetric links. We cannot conclude at a given time named t that disturbances from links B to A and A to B are identical. The left curve in Figure 3.21 shows this aspect of asymmetrical behavior. In fact for symmetric link, the represented RSSI values should follow the diagonal and not dispersed around the diagonal as seen in Figure 3.21. According to our experimental results and those from Heurteufeux *et al.* [7] conducted on large scale experimental sensor nodes platforms (i.e SensLAB), communication links are mostly symmetrical when considering 0 dBm as the transmitting power. This is the case for SensLAB Strasbourg and Lille platforms in the work of Heurteufeux *et al.* [7]. However the links are asymmetric at the Grenoble platform. In addition, works in [7] also show a very high measurement dispersion on the large-scale SensLAB experimental platform. The authors come up with this conclusion: *RSSI over a single link is not always bidirectional, as often assumed. The majority of links seems quite symmetric, but a significant part of the links are clearly not depending on the used emitted power.* In fact a slight proportion of the considered link (about 10 %) can be asymmetric.

RSSI values are log-normal distributed with respect to the distance? This paragraph studies whether RSSI readings are log-normal distributed with respect to the distance. Our aim is to use the RSSI metric to localize a blind mobile wireless sensor node carried by a person or an object in any indoor environment using few wireless emitters nodes. Assuming the RSSI is an indicator function of the distance i.e. a decreasing function of the distance; we first need to select a propagation model in order to map a distance to the attenuation experienced by radio packets and measured by each wireless sensor node. This propagation model should represent

a good compromise between accuracy, feasibility as well as complexity. Indeed, our algorithms have to be performed on embedded devices and hence should not require too much computation or too much data exchange between nodes. The propagation model meeting these criteria and most often used to model indoor and outdoor propagation is the Log-Normal Shadowing Model named LNSM. Such model is described in Chapter 2 of this document. LNSM model has been involved in lots of works [133, 149, 5, 68, 99, 150, 151, 7, 152].

This model relates the path loss called PL(d) between an emitter and a receiver node to the distance that separates these two nodes (d) according to equation (3.2):

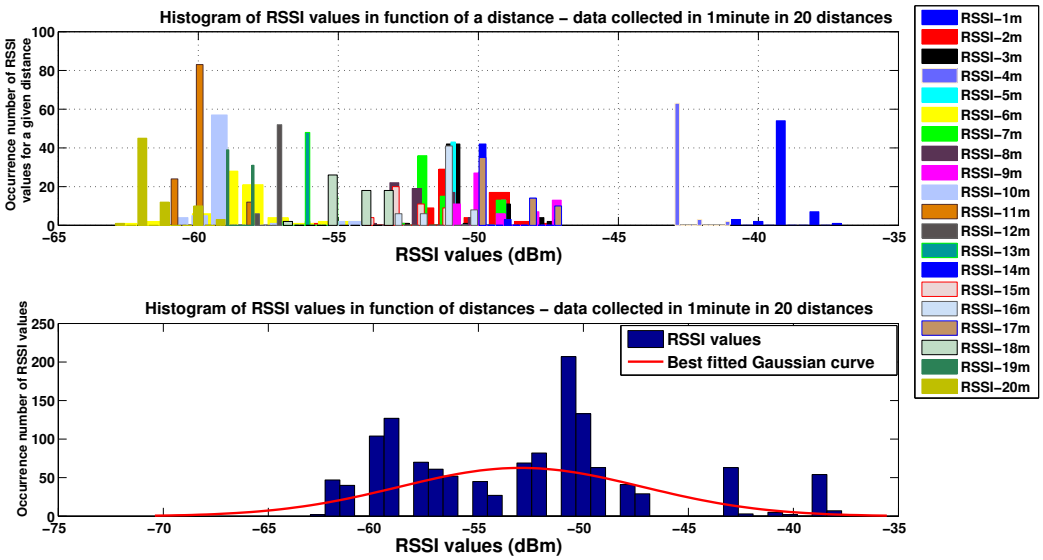
$$\mathrm{PL}(d) = \mathrm{PL}(d_0) + 10.\eta.\log_{10}\left(\frac{d}{d_0}\right) + \mathcal{N}(0,\sigma^2)\,, \tag{3.2}$$

Where the constant calibration offset PL(d_0) is a reference path loss (in dBm) measured at a distance d_0 (usually equal to 1 m in indoor), η is the attenuation factor of the environment, and $\mathcal{N}(0,\sigma^2)$ is a Gaussian random variable with 0 as a mean value and a standard deviation of σ. η reflects the average attenuation of the environment and σ^2 counts for the random effect of shadowing (i.e. large scale fading). This model does not reflect, indeed the effect of fast fading described in Chapter 2 as when considering multiple readings from the same static node, the fast fading effect can be ignored in the LNSM propagation model [5]. The RSSI measured by a wireless receiver is the difference between the transmission power and this path loss. In most of our experiments, as the transmission power is fixed at $0\,dBm$, the RSSI is simply the opposite of the path loss: $RSSI(d) = -\mathrm{PL}(d)$.

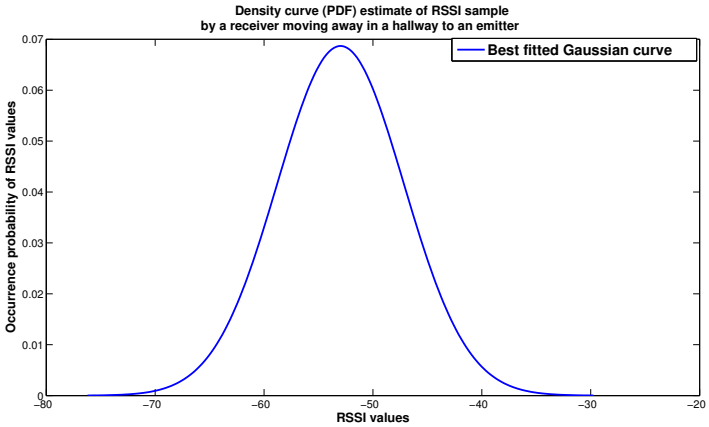
We evaluate in this paragraph whether the measured RSSI values with distance are log-normal distributed. For this, we conducted several measurements in the aforementioned Lincs corridor (cf. Section 3.3.3) composed of one emitter and one receiver node moving away from the emitter over 20 positions starting from 1 m to 15 m with a step of 0.5 m. As the corridor is a waveguide and to reduce the wall reflection effects, the same test was performed in the closed testbed (#1). This test involves one receiver node moving over 40 positions away from one fixed emitter node at $2D$ position (0 m, 0.08 m). Figure 3.22 presents the recorded RSSI values by a receiver node with respect to the distance regarding an emitter node using the *TowardCenter perpendicular* antennas orientation. It shows the presence of noise which is probably due to the shape of the corridor wall (non-straight wall) that creates reflections. However, we still remark the same trend such as the obtained curve with the *TowardCenter perpendicular* architecture which is somehow the more close to the log-normal distribution.

The same experiment performed in the testbed (#1) described in Section 3.3.1, quantifies the RSSI variability and confirms the RSSI distribution over distance. We did not remark any energy hole on measured RSSI values according to a given distance as shown in Figure 3.23. In addition, a linear regression curve from the RSSI distribution is fitted and almost demonstrates a close log-normal distribution with a mean $\mu = -52.976$ and a variance $\sigma = 5.80731$. The patterns observed from experimental results confirm that there are distances information on RSSI values as mentioned in the works of Srinivasan *et al.* [137]. The authors demonstrated that RSSI is not such a bad indicator when its value is above a certain threshold, measured at -87 dBm on the authors' experimental platform composed of IEEE 802.15.4 MicaZ nodes. Unfortunately, this contribution does not provide much detail on the testbed and their conclusion may not hold in a realistic environment such as a typical office. In that case, other results in realistic indoor testbed are needed to certify an absolute log-normal distribution pattern.

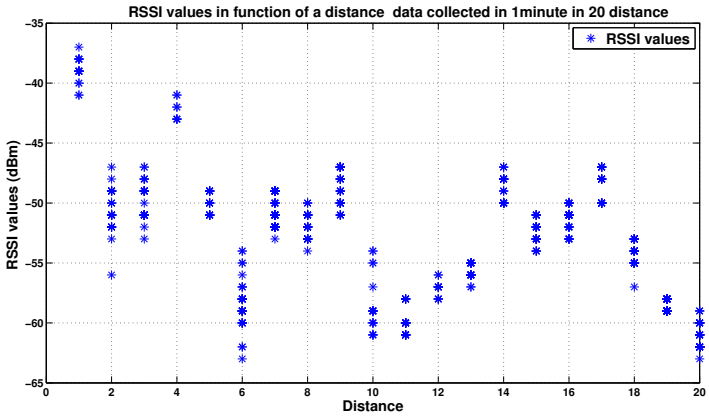
The same experiment as above has been performed in another testbed (#4) (cf. Section 3.3.5).



(a) Histogram of the distribution of RSSI readings over distance



(b) Smooth curve of the best fit Gaussian distribution on collected data



(c) RSSI distribution over distance

Figure 3.22: RSSI distribution between one emitter and one receiver node moving away in the Lincs Isup hallway testbed

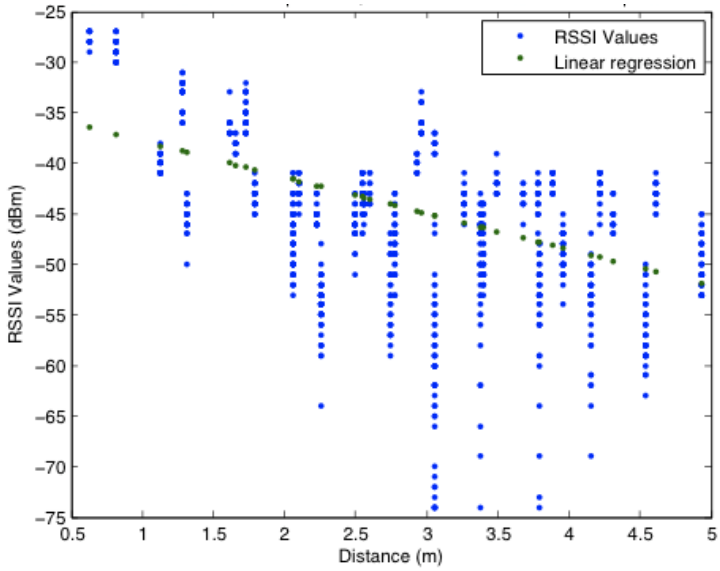
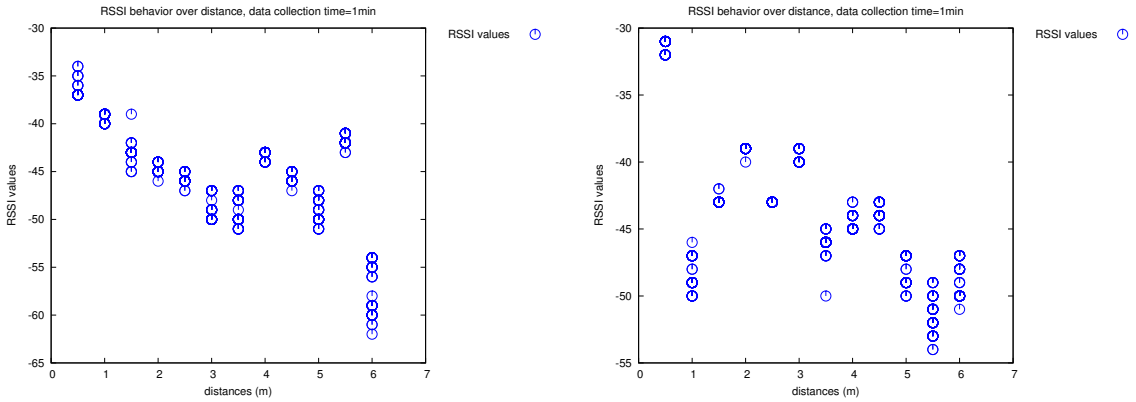


Figure 3.23: RSSI distribution between one emitter and one receiver node moving away over 40 positions in testbed (#1) (cf. Section 3.3.1)

RSSI values between one emitter and a receiver node are measured with two types of configuration. One configuration involves an emitter placed near the furniture and a receiver moving away from the emitter towards the back of the office room (i.e the wall). The second configuration involves a receiver situated in the back of the office room precisely near the wall and the receiver is moving from the wall to the center of the office room, i.e. near furniture. There are one person moving inside the office room and at least two people sitting at their desks. The experiments results using the architecture of the *TowardCenter perpendicular* antennas for these two configurations are shown in Figure 3.24.

The analysis of these two curves are presented in Figure 3.24. Obtained results do not show the same dispersion of RSSI values for the same separation distance between an emitter and a receiver. It is clear that the furniture has disruptive effects on signal. A shape nearly to the normal distribution is reported with the left figure, which is not the case for the right one. Consequently, depending on how emitter and receiver are oriented with respect or not to the furniture, we do not report the same behavior concerning RSSI values for the same distance. This suggests and confirms a need to take into account in the localization algorithm the environmental orientation of the landmarks and of the receiving target node antennas.

To confirm the above results, we draw the histogram of the collected RSSI values. We also draw the histogram of the same RSSI samples in function of the distance for both different configurations of emitter-receiver node deployment. The smooth curve of the best fitted gaussian distribution based on the same data samples is also drawn using Matlab for both of the two deployment configurations of emitter-receiver node (cf. Figures 3.25). Histograms related to the configuration in which the emitter is near the furniture and the receiver is moving to the wall are presented in Figure 3.25. Such histograms when compared to the best fitted Gaussian curve show a Gaussian distribution of the RSSI samples with mean $\mu = -45.4025dBm$ and variance $\sigma = 5.2516$. The achieved results confirm our hypothesis that is the layout of the furniture in indoor environment plays an important role (i.e. disruptive or not) in signal propagation effects.



(a) TX near the furniture - RX moving to the wall (b) TX near the wall - RX moving to the furniture

Figure 3.24: RSSI behavior depending on the orientation of emitter (TX)-receiver node (RX) moving away with respect or not to the furniture in testbed (#4) (cf. Section 3.3.5)

Table 3.1: Known reference path loss PL(d_0) and path loss exponent values η from 5 landmarks named TX in testbed (#1)

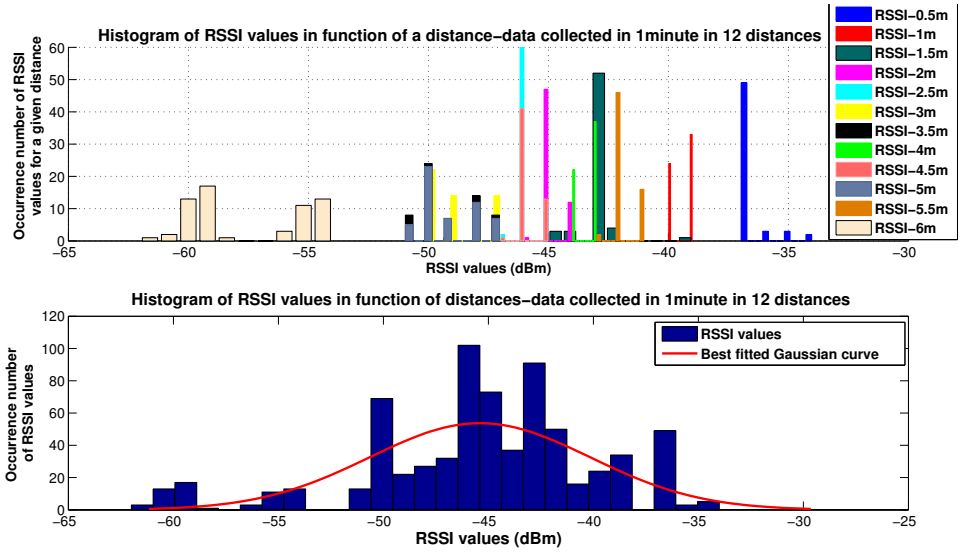
	TX1	TX2	TX3	TX4	TX5
PL(d_0)	41.5016	28.9089	42.9325	56.5850	39.5202
η	1.8462	3.5697	2.3185	1.8165	1.6783

Influence of Multipath on RSSI Figure 3.26 which uses the same deployment scenario as in Figure 3.24 (cf. testbed (#4) in Section 3.3.5) shows the variation of the RSSI values collected by two equidistant receivers situated at 5.6 m from the emitter. The distributions of the RSSI values for each receiver regarding the emitter are not the same even if the distances between the receivers are equal. In fact, one of the receivers has a misbehavior which results from the proximity with the indoor obstructing factors (furniture, obstructing object, layout of the room).

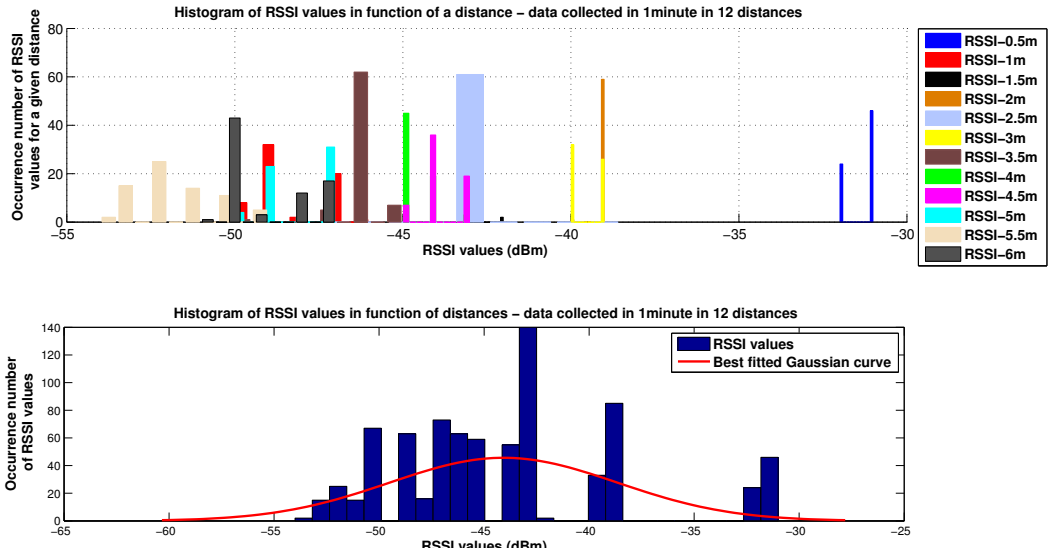
Figure 3.27 on the other hand, performed in testbed (#3) represents, for a given position, the distribution of the RSSI values from each landmark and the number of successfully received beacons.

Experimental results emphasize with this example that the node whose values are shown in the second position from the top is a bit far in terms of dispersion of RSSI values compared to the others. Those results are a key idea of considering that a node can be affected by an abnormal comportment. In other word, in environment with no or little perturbation we should measure equal or lower variations on the standard deviation because the environment doesn't change and also the propagation condition remains the same. Hence, the idea to consider the propagation model parameters dependent on landmark location as shown in Tables 3.1 3.3. Such results are first published in one of our works [153] in which the linear regression was used to compute the values of the parameters η and PL(d_0) when considering a minimum mean square error (MMSE) of the sensed data.

In fact, the same experiment has been performed in in our laboratory, in different mentioned testbeds to clearly advocate that the dominant factor is the multipath. Recall that multipath are introduced by reflection, diffraction phenomena or a combination of them. It arises from obstruction of the signal happening along the communication path. One should note that changes appearing on a wireless node in reception mode vary differently depending on the data source



(a) TX near the furniture - RX moving to the wall



(b) TX near the wall - RX moving to the furniture

Figure 3.25: Histograms of RSSI readings distribution depending on the orientation of emitter (TX)-receiver (RX) node moving away with respect or not to the furniture in testbed (#4) (cf. Section 3.3.5)

Table 3.2: Known reference path loss PL(d_0) and path loss exponent values η from 4 emitters named TX in testbed (#2)

	TX1	TX2	TX3	TX4
PL(d_0)	54.6315	52.8471	28.5622	53.0334
η	1.0143	1.4663	3.3471	2.4317

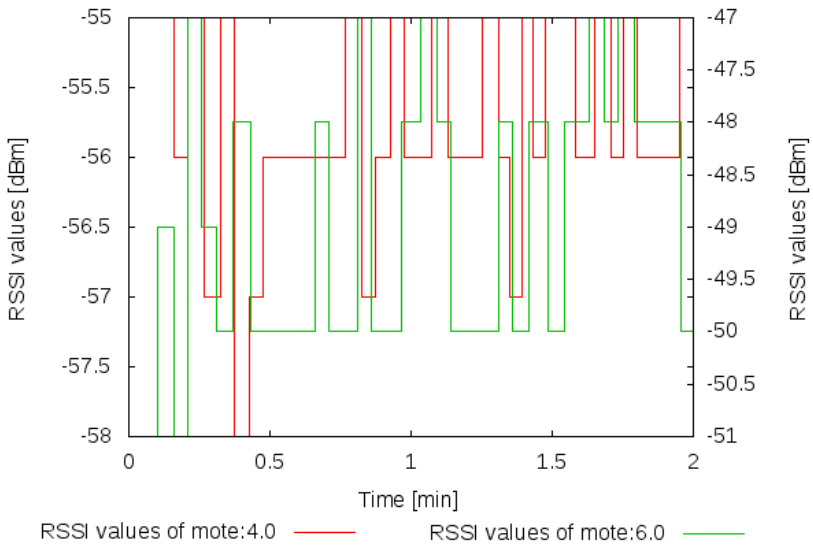


Figure 3.26: Overview of the multipath effects on RSSI values in indoor environment: RSSI in function of the distance on two equidistant receivers situated at 5.6 m from the emitter

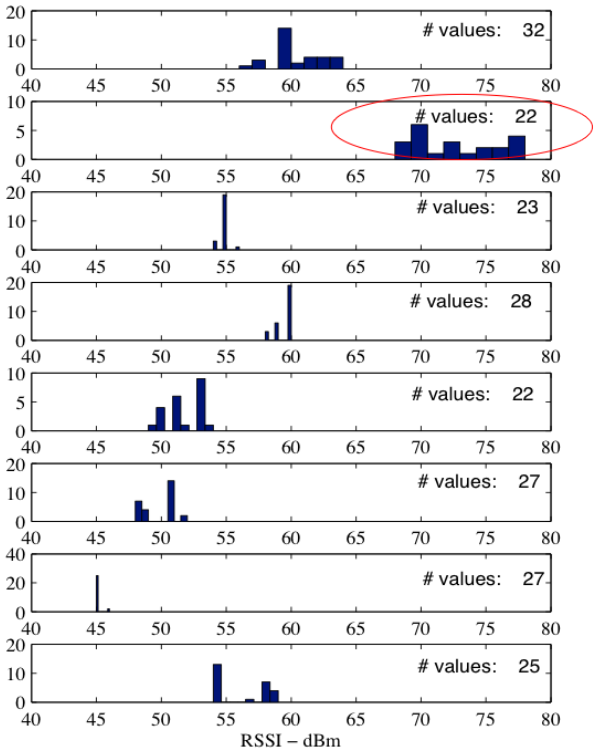


Figure 3.27: An histogram of RSSI distribution from each landmark at testbed (#3)(cf. Section 3.3.2)

Table 3.3: Known reference path loss PL(d_0) and path loss exponent values η from 8 landmarks named TX in testbed (#3)

	TX1	TX2	TX3	TX4	TX5
PL(d_0)	50.3452	49.1444	55.3127	40.4212	44.9725
η	1.0950	0.8237	0.4307	2.0919	1.2565
	TX6	TX7	TX8		
PL(d_0)	41.7087	51.8631	61.1757		
η	2.1450	0.7759	0.6272		

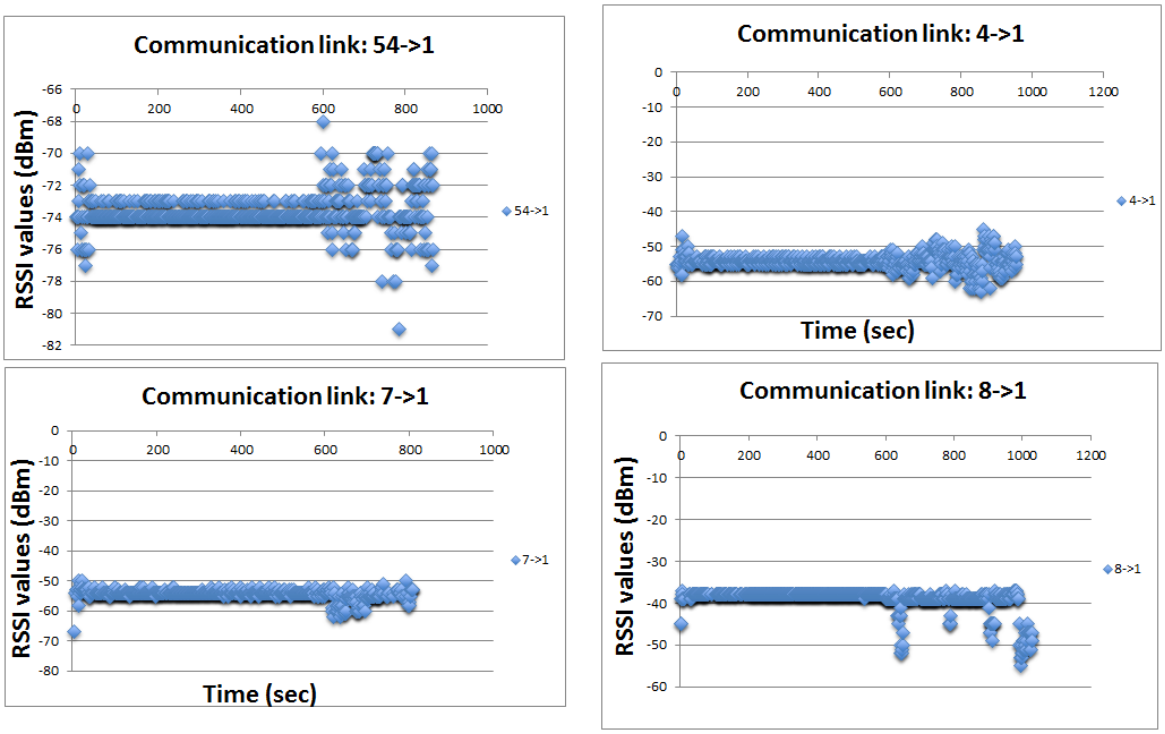


Figure 3.28: Pattern of measured RSSI values on receivers nodes observed for a particular emitter node (use case 1)

named emitter node. This means that the RSSI values measured from an emitter node may be affected by fluctuations and these are even more important when the emitter node is in less favorable propagation conditions than another emitter node. This is highlighted in the graphs in Figure 3.28 in which we observe a pattern for a given receiver node (named node 1.0): when there is a change on a link, this can be found more consistently on another link. As an example the signal from emitter node identifier 7.0 to receiver node identifier 1.0 is reliable up to $t = 600sec$ then it is affected by slight variations up to a gap of 5 dBm. Instead, the signal from the emitter node 8.0 to the receiver node 1.0 is stable up to $t = 600sec$ and then modified by high variations up to 10 dBm difference. Such result is also verified for the link 4.0 to 1.0 and for the link 54.0 to 1.0 as well in which we notice a gap of 8 dBm and a curve with a little more fluctuation.

Obviously the receiver node 8.0 is not well situated in terms of propagation condition; this is the case in the experimental platform in which this node is very close to a back wall of the testbed. There is several fluctuations in reception mode for this node whatever the emitted node is (e.g. link 7.0 to 8.0, link 4.0 to 8.0 and link 54.0 to 8.0) as shown in the graphs in Figure 3.29.

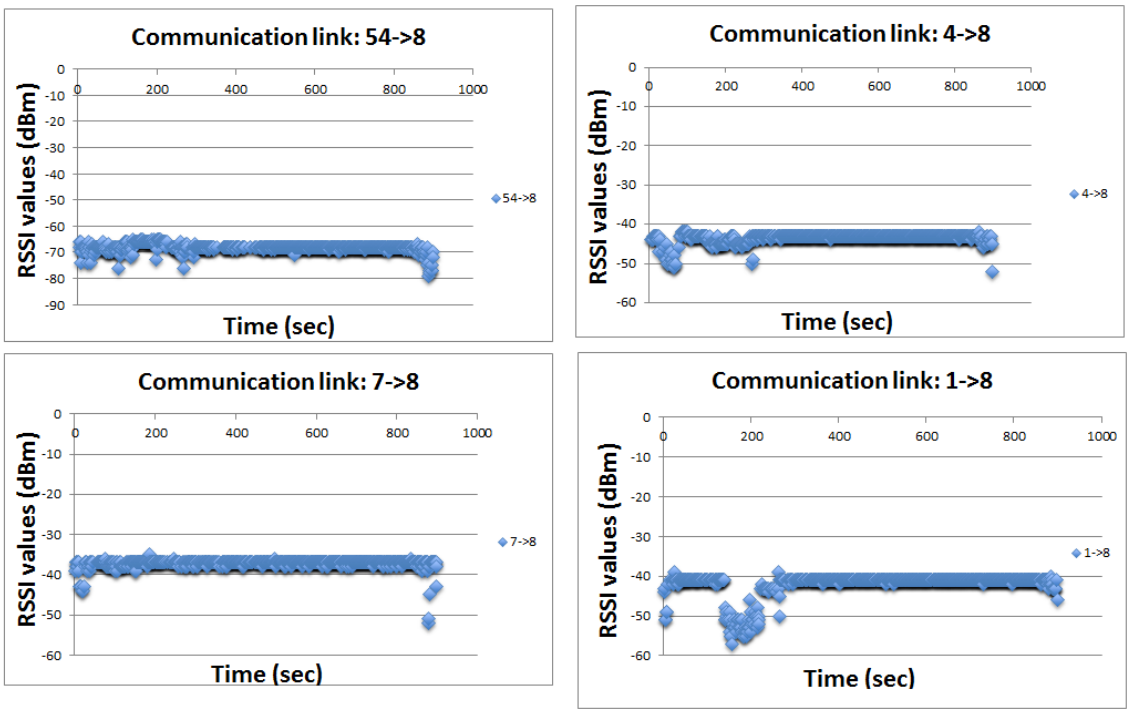
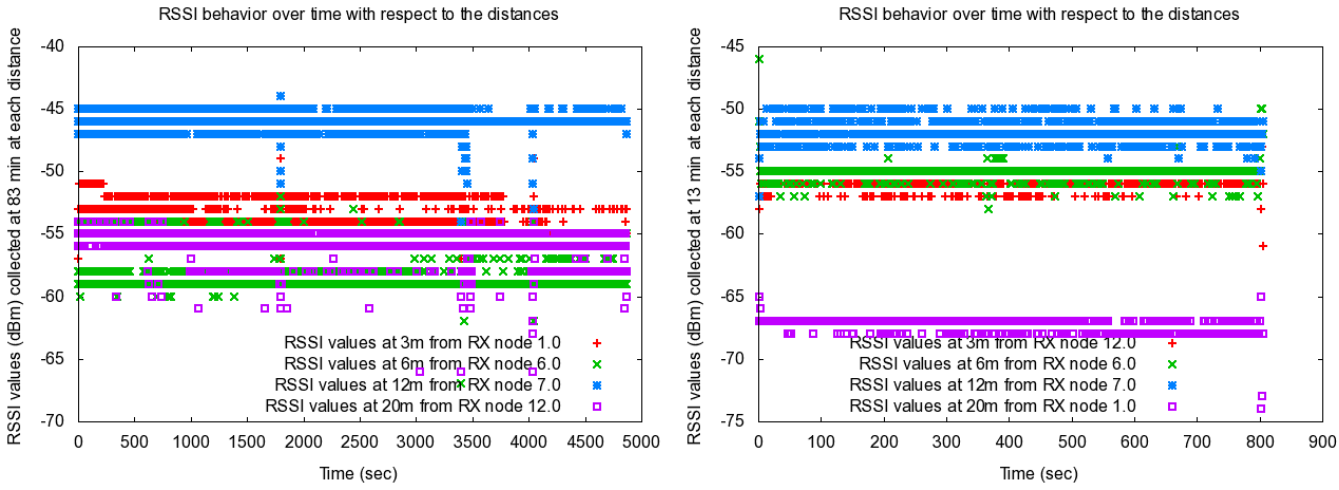


Figure 3.29: Pattern of measured RSSI values on receivers nodes observed for a particular emitter node (use case 2)

Furthermore this variation is even more important if the emitter node is not well located due to the presence of disturbing factors in the communication link. The curve of the link 7.0 to 8.0 is much smoother and stable compared to the curve of the communication link of 54.0 to 8.0 and even smoother and presents less variation than the curve of the communication link 4.0 to 8.0.

Quality of the circuits of sensor nodes (hardware quality) We evaluated in this paragraph the quality of the circuits of sensor nodes by measuring the RSSI between an emitter and four different receivers situated respectively at 3 m, 6 m, 12 m and 20 m in Lincs isup corridor testbed (cf. Section 3.3.3). In ideal case which means no change coming from receiver sensitivity neither from environmental disruption, we should measure decaying RSSI values when moving from one distance to another greater one. We additionally should observe constant RSSI values for a given distance. Instead from the reported results in the experiment 1 which results are presented in Figure 3.30, the measured RSSI values at a distance of 3 m by a node named 1.0 are greater than those at the distance of 6 m measured by a node 6.0. In experiment 2 (cf. Figure 3.30), reported RSSI values at the distance of 3 m at the node 12.0 are almost equal to those reported at the distance of 6 m from the node 6.0. In addition at the distance of 12 m and 20 m in the experiment 1 (cf. Figure 3.30), the reported RSSI values are not decaying with the distance if we take into account the dispersion of RSSI values from all involving distances that are at 3 m, 6 m, 12 m and 20 m. However the measured RSSI values at 12 m on node 7.0 are greater than those measured at the distance of 20 m and also greater than the RSSI values reported at the distance of 6 m.

Clearly in the experiment 1 in Figure 3.30, the node located at the distance of 12 m and named 7.0 and the node being at the distance of 20 m and named 12.0 have abnormal behavior because of their sensitivity. The experiment 2 in Figure 3.30 however reports smaller RSSI readings



(a) Experiment 1

(b) Experiment 2

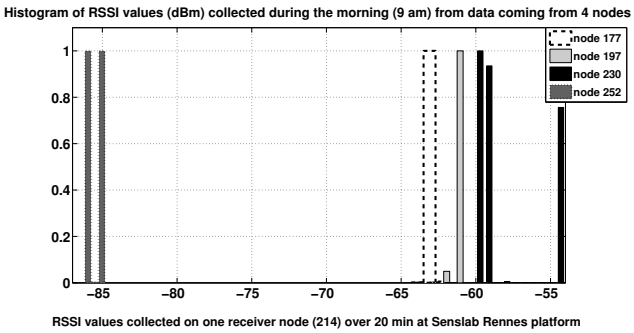
Figure 3.30: RSSI behavior over time with respect to the distance

at the distance of 20 m reported by the node 1.0 compared to RSSI readings reported at the distance of 12 m by the node 7.0. In fact at the same distance of 12 m, the same node named 7.0 involved in the same repeated experiment in two separate days in the same environment and conditions show the RSSI values dispersed in the interval of $[-50 - 45]\,dBm$ for the first experiment and between $[-55 - 50]\,dBm$ for the second experiment. These results demonstrated a receiver sensitivity highlighted by a gap of 5 dBm between the two intervals of the RSSI values for the two experiments.

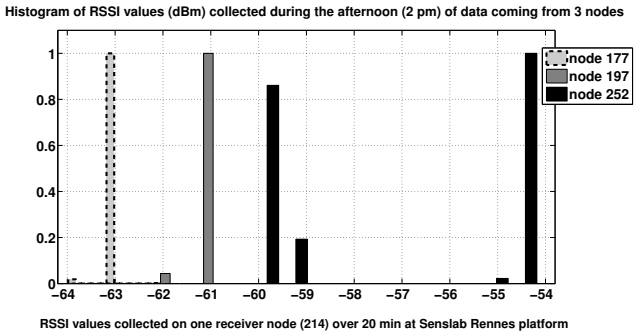
In addition, results reported by both nodes 12.0 and 1.0 in both Figures emphasize the disruptive effect of receiver sensitivity through different radio chips. We noticed that the RSSI values measured at two different receivers (nodes identifiers 12.0 and 1.0) are different even though they were measured at the same distance and in similar environmental conditions: all parameters that affect the received signal strength are constant. On the node with identifier 1.0, at the same distance of 20 m in the experiment 2 in Figure 3.30, the RSSI readings are around $[-67 - 65]\,dBm$. Comparing these values to those measured at the distance of 3 m, 6 m and 12 m, the principle that the attenuation decreases with the distance is validated. This is not the case with the experiment 1 in Figure 3.30 in which we found the RSSI values reported at the node 12.0 at the distance of 20 m being between $[-63 - 54]\,dBm$, thus greater (i.e. attenuation is smaller) than the RSSI values measured at the distance of 6 m by the node 6.0.

Is indoor radio channel varying with the time? A question solves in this paragraph is the following: does the time of day have an influence on the measured RSSI values? does the start time of the measure have an influence? From how many samples can we trust the measured RSSI values? To answer these questions we performed the same experiment at different times of the day with five nodes including four emitters and one receiver in SensLAB Rennes platform (cf. Section 3.3.4). A receiver node measures RSSI values from four emitter nodes during 20 minutes in the morning at 9 am, in the afternoon at 2 pm and in the evening at 9 pm. Figure 3.31 shows the achieved experimental results.

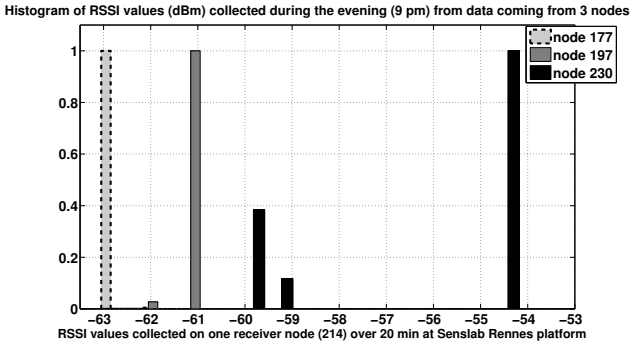
We notice that the RSSI values reported during the morning experiment by the node with identifier 214 from the emitter node with identifier 177 are distributed between −62 dBm and



(a) Histogram of RSSI values (dBm) collected in the morning at 9 am



(b) Histogram of RSSI values (dBm) collected in the afternoon at 2 pm



(c) Histogram of RSSI values (dBm) collected in the evening at 9 pm

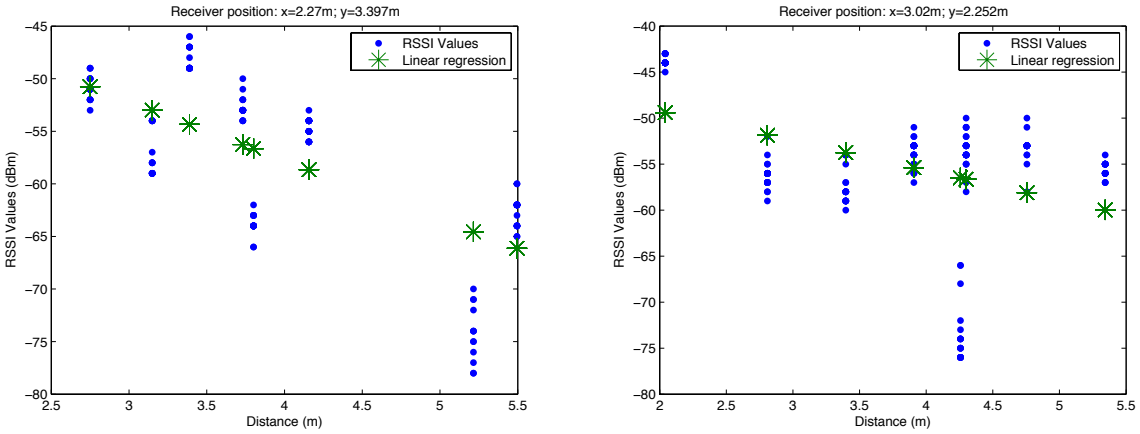
Figure 3.31: Histogram of RSSI values (dBm) collected during the morning, the afternoon and the evening

−63 dBm. These values moved slowly between −64 dBm and −63 dBm during the afternoon and remain stable and same for the evening. In addition, the value −63 dBm appears more than the others (almost 100%). RSSI values measured at the same node with identifier 214 from the message coming from the node with identifier 197 are spread between −62 dBm and −61 dBm in the morning while the same values are measured during the afternoon and the evening. However the measured RSSI values recorded at the link 230 → 214 are observed between −59 dBm, −58 dBM, −57 dBm and −54 dBm in the morning, −60 dBm, −59 dBm, −55 dBm and −54 dBm in the afternoon and stable between −60 dBm, −59 dBm and −54 dBm in the evening. The proportion of values from −60 dBm and −54 dBm are dominated for this communication link. Finally RSSI values measured for the data packets coming from the node with identifier 252 are only observed during the morning and distributed in these two values −86 dBm and −85 dBm with the same proportion. Because of environmental disturbance, this node is somehow far from the receiver node with identifier 214.

To summarize, the distribution of the RSSI values do not change as much during the time of the day unless with the presence of factors that can disturb the signal propagation between an emitter and a receiver such as people walking around, the presence of an object in the communication link like a chair, a table, a cupboard.... Notice that the receiver with identifier 214 has received data from the emitter node with identifier 252 only from the morning experiment. The rest of the time it does not measure RSSI values from the emitter node with identifier 252 which is probably in unfavorable propagation conditions with respect to the communication link of the receiver node 214. In conclusion measuring the RSSI values in indoor environment at different time of the day in the same propagation conditions should not have a big impact on localization accuracy. However, the indoor environment dynamics and changing behavior prevent to consider the parameters of the propagation model constant regarding any indoor environment. Such phenomena also prevent using the same propagation model parameters every time there is a need to localize a target in a nearest indoor environment. For instance, in [3] with the use of the log-normal shadowing model, path loss from different indoor environment are measured and then averaged. Measured path loss is then kept constant and used to localize any pedestrian in a warehouse. But, in the case the propagation of the indoor environment changes (e.g. presence of walking people), the constant estimated path loss can deviate from the previous estimation and thus affects the localization error that should be perceived. Our suggested solution is to renew the estimation and in a dynamic way the values of the propagation model parameters for any indoor environment.

3.4.4 RSSI versus distance

In the following we additionally describe how data are distributed when distance increases in different testbeds. We use two distinct testbeds with sizes: testbed (#1) and (#3). Figures 3.32(a) and 3.32(b) represent two sets of distributions of the measured RSSI values by a blind mobile node in function of the distance from beacons received from 8 different emitters in two representative positions. Both positions are randomly picked in testbed (#3). The blue dots in Figures 3.32(a) and 3.32(b) reproduce measured RSSI values, while green stars represent the result from a linear regression interpolation performed on the RSSI sample. In most cases, the RSSI values are only affected by noise, and the variance of the RSSI remains acceptable.
However, for a few communication links between the mobile node and each emitter (seventh scatter in Figure 3.32(a) or fifth scatter in Figure 3.32(b)), the variance becomes very large, which introduces a bias. As these large variations do not always concern the same emitter, they are due to multipath propagation effects and shadowing (cf. Figure 2.1 in Chapter 2).



(a) 1st testing position (#11): x=2.27 m; y=3.397 m (b) 2nd testing position (#17): x=3.02m; y=2.252 m

Figure 3.32: RSSI as a function of the distance measured at one receiver node from the 8 landmarks in two arbitrary positions in testbed (#3)

To confirm/validate the results observed above we conduct the same experiment in testbed (#1). We use statistics box plots tool to compare RSSI values in connection with distance. Measured RSSI samples are gathered in testbed (#1) over 48 different positions of a mobile receiver node. We randomly choose 5 of the 48 mobile receiver node positions. We then draw in Figure 3.33 as box plots, the attenuation experienced with each of the 5 emitters at each of these 5 positions. Note that a box plot or whisker diagram is a standard statistical tool of representation of the distribution of data. Box plot representation is useful in detecting either shifts and large variations on data sets or the shape of the corresponding data set. In case where the data set contains abrupt change for outliers, they are plotted separately outside the chart of the box. As presented in Figure 3.33, a box plot displays distribution of data according to the data set lowest value (minimum), the size of the first quartile, the median value, the size of the third quartile and the data set highest value (maximum). The central rectangle of the box-plot is delimited by the first quartile to the third quartile. The segment inside the rectangle is the median value and those small and outside the rectangle, precisely those segments situated above and below the box are named "whiskers". The positions of the lowest and highest value are shown by whiskers. A box-plot is complement to a histogram and a helpful tool to formulate the median value and to characterize any distribution. In the following we use box plots to perform simultaneous comparisons in order to figure out how the RSSI values are distributed in relation to distance.

In ideal case, as long as the median is high, when moving toward greater distances, box plots of corresponding measured RSSI values should show effective Gaussian shape. This Gaussian shape is remarked for the first two distances of the mobile receiver (i.e $distance = 1.69$ m and $distance = 1.88$ m) in which RSSI values are measured on data sent by the fifth emitter. The same result is also observed for the first two distances from the fourth emitter. However, we do not notice the same results for all distances from the emitter 1 and also for the three first distances separation between the mobile receiver and the third emitter. The height of the box represents the dispersion of the measured RSSI values and a small box means all measurements lie in a small range of values, in other word a small dispersion. The red crosses (+) indicate the presence of outliers on the measured RSSI values, this mean these measured values are too dispersed to be correlated to the others. From the results observed on the fifth emitter,

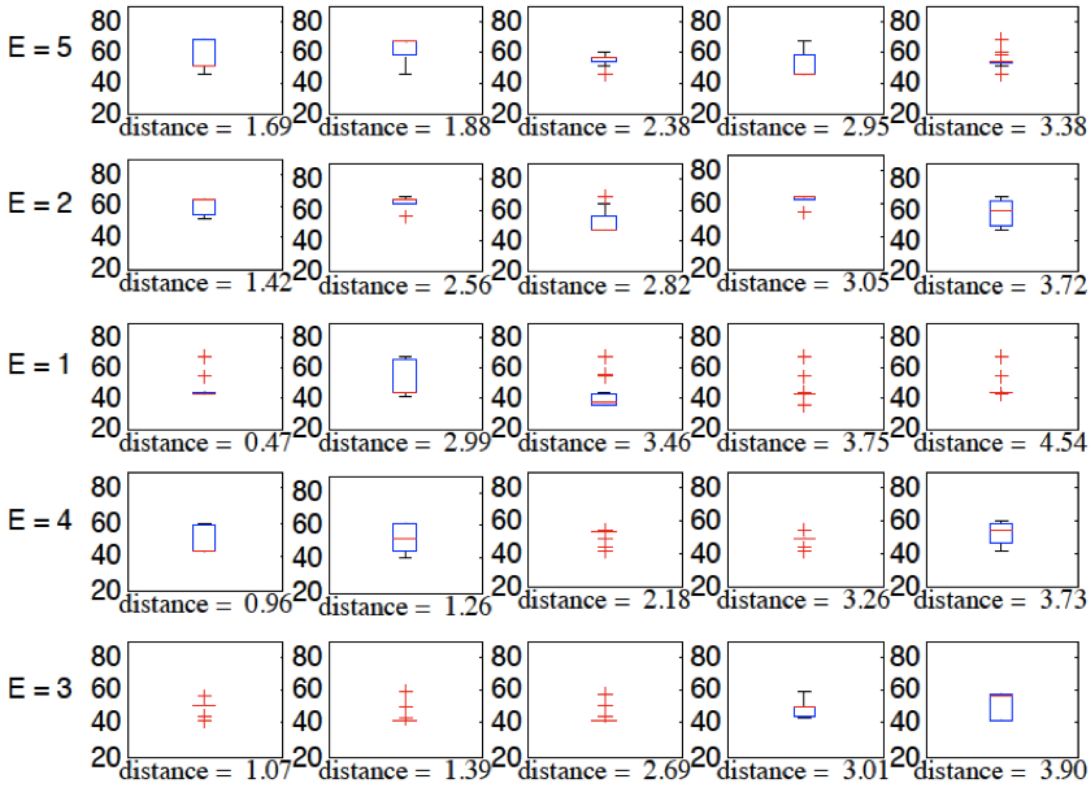
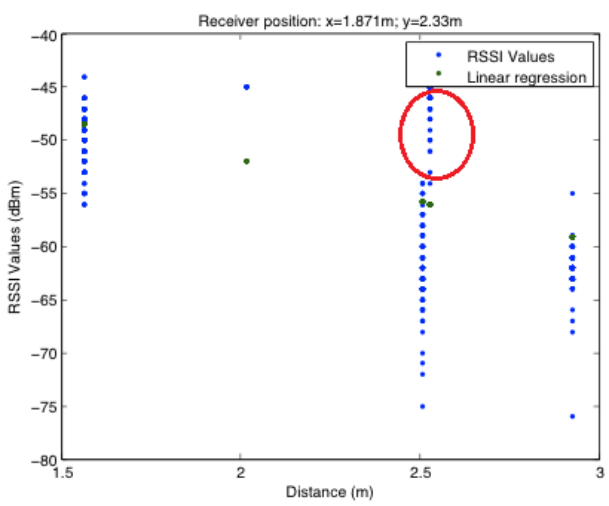


Figure 3.33: Boxplots of the data collected on the receiver at 5 random distances in meter from each of the 5 emitters in the testbed (#1)

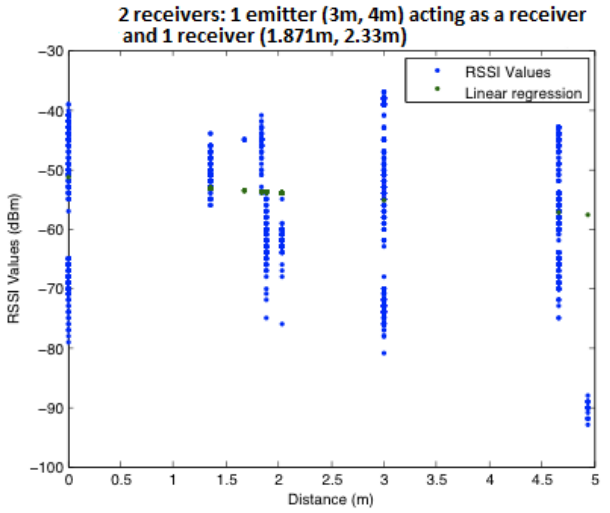
a highest signal strength is reported at *distance* = 1.88 m. The same trend is reported at distance *distance* = 3.05 m. Such results from box plots confirm that RSSI values are not always correlated with distance. Due to environment disturbances that create multipath, the log-normal shape of RSSI readings is not always respected when the mobile receiver evolves toward greater distances.

Furthermore, we represent the distribution of the RSSI values by selecting two random positions of the receiver and we measure during 1 minute interval the RSSI values of data packets coming from the 5 emitters in testbed (#1). Packets were broadcasted randomly by each emitter at each second. The RSSI distribution of collected data from the 5 emitter nodes and measured at one receiver node position (i.e. left curve in Figure 3.34) as well as at two different receivers positions (i.e. right curve in Figure 3.34) are illustrated in Figure 3.34. The right curve in Figure 3.34 represents the measured RSSI values at one receiver position while the right curve in Figure 3.34 is obtained by measuring the RSSI values at two distinct receivers positions. Such experimental results highlight that if the RSSI values of the emitters are not learned separately at the receiver node, we would not verify the hypothesis that the RSSI values decrease as a function of the distances according to log normal shadowing propagation model in indoor environment.

From these plots, we note that unless we distinguish RSSI values from emitters by treating data that comes from each of them separately, RSSI measurements at a single receiver are not clearly correlated with distances. In fact, reference path loss parameter and path loss exponent are different for each emitter. Table 3.4 reports estimations of the reference path loss PL(d_0) and path loss exponent η parameters calculated using maximum likelihood estimation over the full measurements set and for all possible receiver positions in a given testbed.



(a) 1st testing position (#29): x=1.871 m; y=2.33 m



(b) 2nd two testing positions (#29): x=1.871 m; y=2.33 m and x=3 m; y=4 m (1 landmark)

Figure 3.34: RSSI as a function of the distance measured by one (i.e. left curve) and two receivers in two arbitrary positions (i.e. right curve) from 5 landmarks in testbed (#1)

Table 3.4: Known reference path loss PL(d_0) and path loss exponent values η from 5 emitters named TX in testbed (#1)

	TX1	TX2	TX3	TX4	TX5
PL(d_0)	41.5647	40.0314	41.2408	58.2899	40.9747
η	1.8058	1.3083	2.7786	1.5726	1.3893

Results shown in Table 3.4 above is in conformity with those of [99] about the variability of the path loss in indoor, but we also demonstrate that the reference path loss PL(d_0) is variable in addition to the path loss exponent η. Results reported in Figure 3.34 recall again to learn each emitter separately otherwise there will not be any correlation between measured RSSI values and the distances. In addition, in Figure 3.35 we can denote that attenuation is different depending on the location of an emitter. It can also be seen on the set of values from Figure 3.35, a behavior that matches the log normal shadowing model pattern: RSSI decreases as a function of distance.

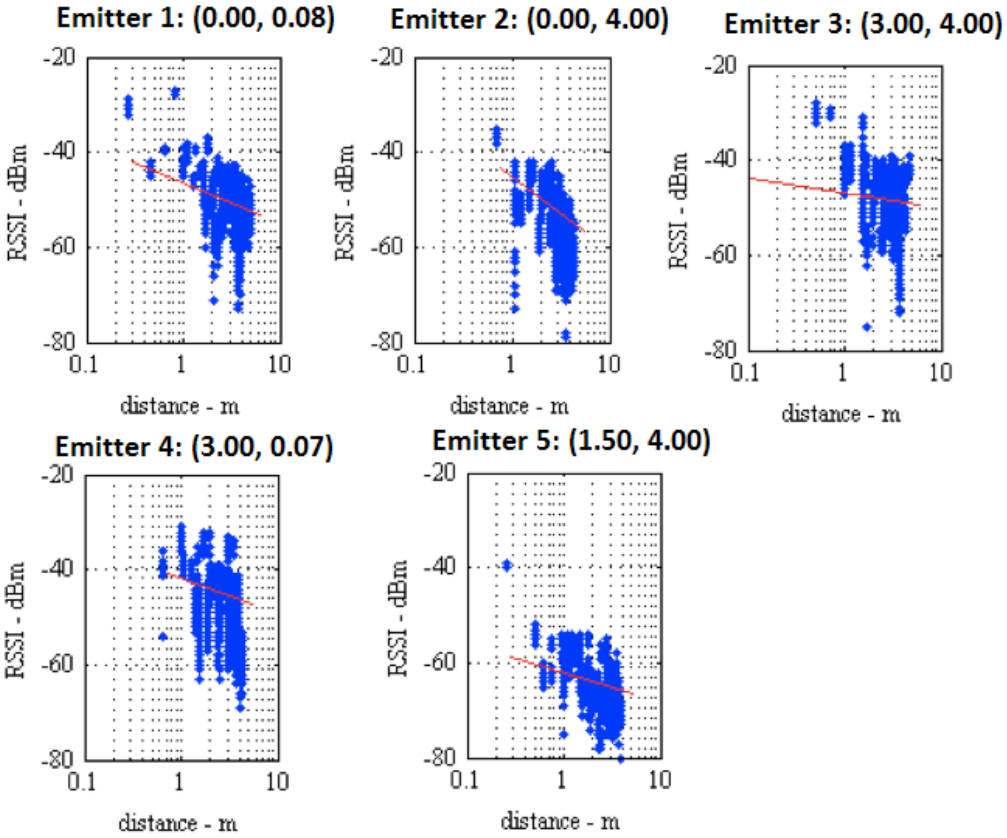
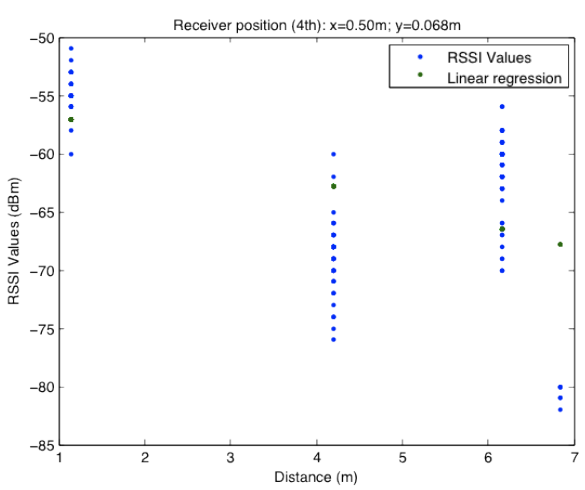


Figure 3.35: RSSI (in dBm) versus distance (m) at the receiver collected from each emitter in testbed (#1); red lines are linear regression prediction

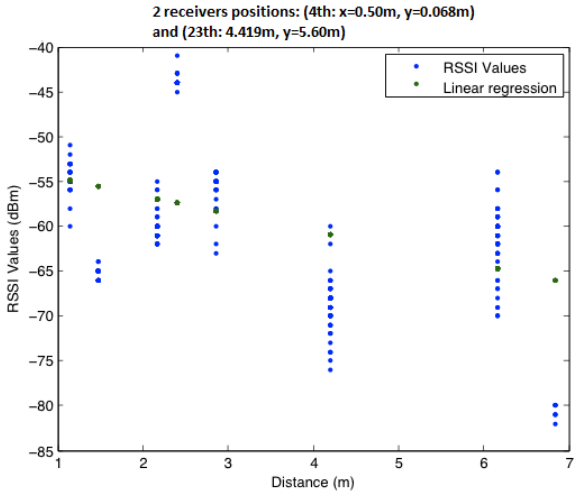
However, the dispersion of values is still very important and cannot lead to a good accuracy or a small confidence interval on the prediction of the distance calculated from the maximum likelihood. For example in Figure 3.35 for the points that are 4 m away from the emitter 1 with $2D$ coordinates: 0.00 m, 0.08 m, there are variations of RSSI values ranging from -65 dBm to -40 dBm, so almost a gap of 25 dBm, which is very wide. In addition, we observe that the attenuation depends on the environmental locations of both receivers and emitters which are different from one corner to another.

Another experiment is realized in testbed (#2). Among these 5 nodes, 4 are emitters and one is a receiver node. We choose two random positions of the receiver to measure the RSSI distribution in function of the distance. Packets were broadcasted randomly by emitters between 2 and 4 seconds. After a successful testbed setup, RSSI values reported at one and then two receivers positions are then plotted according to all emitters and presented in Figure 3.36.

Table 3.5 reports estimation of the reference path losses PL(d_0) and path loss exponents η



(a) 1st testing position (#4): x=0.50 m; y=0.068 m



(b) 2nd two testing positions (#4): x=0.50 m; y=0.068 m and (#23): x=4.419 m; y=5.60 m

Figure 3.36: RSSI as a function of the distance measured by one (i.e. left curve) and two receivers in two arbitrary positions (i.e. right curve) from 4 emitters in testbed (#2)

Table 3.5: Known reference path loss PL(d_0) and path loss exponent values η from 4 emitters named TX in testbed (#2)

	TX1	TX2	TX3	TX4
PL(d_0)	54.5084	52.1867	36.8205	47.9144
η	1.3544	1.5651	2.4060	3.0286

parameters computed using maximum likelihood over the full measurements set and for all possible receiver positions in the testbed (#2).

Results from Figure 3.36 and Table 3.5 confirm previous experiments results: RSSI values should be considered separately related to each emitter. As the average values of the RSSI samples measured at different positions do not always decrease as a function of the distance, we confirm that the parameters of the propagation model at least differ from one emitter to one another, which prevents from treating all emitter nodes equally in a statistical estimation. A localization algorithm should therefore takes into account the possibility that the signals obtained from one emitter may be affected by such effects and should not be taken into account in the localization process. We also verify from the blind mobile receiver node the non-presence of lots of outliers or abrupt change or even large gap between the measured RSSI values, only very few outliers were remarked. Hence we consider a node among the others in a network can be affected by abnormal behavior due to the multiple paths and indoor signal perturbation factors. Then the effects of corrupted data sample coming from that abnormal node are reduced following a probability distribution (cf. Chapter 4).

Characterization of measured RSSI variation and non-conformity of data smoothing methods for filtering Figures 3.37 and 3.39 introduce experiment results obtained from three different testbeds. We simultaneously measure on 3 receivers nodes the RSSI values sent by an emitter node located in testbed (#5) (cf. paragraph 3.3.5). The two first receivers nodes are located in the same testbed (#5) as the emitter node: the first receiver is distant to 3.65 m

from the emitter node and the second one is separated from 7.3 m of the emitter node across a desk. The third receiver node which is 18.69 m away from the emitter node, is situated in a meeting room of size 7.20 m width, 7.12 m long and separated to the emitter node by the Lincs isup corridor testbed. The meeting room is often used by about fifteen people.

Since the RSSI values are mainly affected by noise, we propose to smooth such noisy data measured on a receiver node using Exponentially Weighted Moving Average (EWMA) statistics methodology [154, 155]. EWMA is always used to detect small shift through changes accumulated under exponential smoothing in a sequence of independent samples [156]. EMWA technique is not sensitive to normality assumptions [157]. The EWMA technique gives greater weight to recent measurements samples as its value is regularly computed. This shows the effect of process improvements. The EWMA approach introduces the weighted parameter α called the smoothing factor/parameter. Under the condition that $0 < \alpha \leq 1$, the weight parameter α is viewed as the way to determine how the prior data are smoothed. To this aim, when the value of α is close to 1, the dampening of the data is quick and when the value of α is close to 0, the dampening is slow. In other word, the value of α parameter should be selected depending on the one that gives the smallest mean squared error (i.e MSE) and that best smooth the original curve. The choice of smoothing parameter α is carefully chosen to make the chart more or less sensitive to these fluctuations.

According to our experimental results the weight parameter $\alpha = 0.01$ is the one that gives better results compared to the other tested weight parameters in the interval $[0; 1]$. IIn fact, we first measure the RSSI values on a receiver node from data sent by the three aforementioned emitter nodes that are respectively distant from 18.69 m, 7.3 m and 3.65 m to the receiver node. We then apply the EWMA approach to filter the measured RSSI values while estimating the corresponding distance using the log-normal shadowing model to validate the filtered data. To this aim the measured RSSI data are filtered in accordance with a reference distance, a reference path loss computed according to the Free Space Path Loss formula presented in equation (2.2) in Chapter 2. The estimated distance is then computed according to the log-normal shadowing model presented in Section 3.4; the filtered RSSI values as well as the corresponding variance σ^2 based on EWMA. At the end, the value of the weighted parameter α is the one that gives a better estimate of the distance that separates an emitter node to a receiver node in line with the true distance separation. The variance computed within EWMA formula is expressed as:

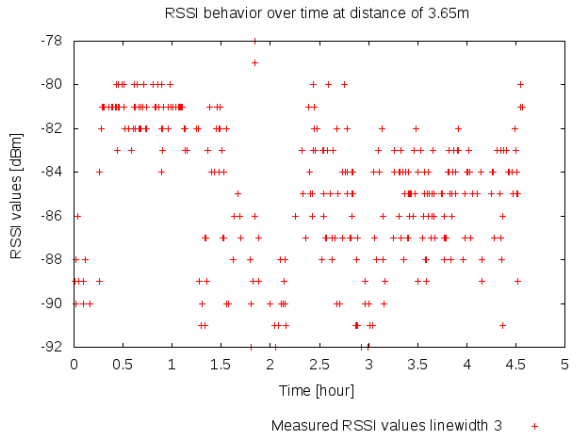
$$variance\_ewma_n = (\alpha)\frac{(rssiValue - mean\_RSSI)^2}{total\_number\_RSSI\_values} + (1-\alpha)variance\_ewma_{n-1}$$

The used algorithm is shown below:

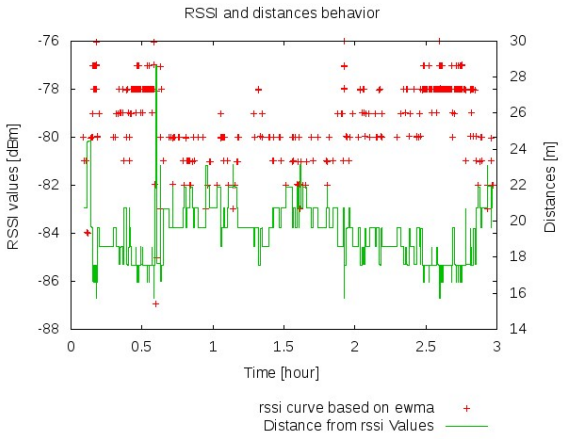
```
#The smoothed RSSI and variance computation using statistics EMWA technique
if ($i ≤ $RSSI_values_number) {
        ($timeValue, $rssiValue) = split(//, $line);

        if ($i == 0) { #The first line of the file containing the measured
                       #RSSI over time at any distance
                $ewma_prev = $rssiValue;
                $variance_ewma_prev = $α2 * (($rssiValue−$mean_RSSI)^2 / $RSSI_values_number) + $Δ
                $Δ = ((1 − $α2) * $initial_variance);

                $smoothed_RSSI = $ewma_prev;
```

```
                $σ = $variance_ewma_prev;
                        $PL(d) = −($smoothed_RSSI);
                }
                #After the first line of the file containing the measured
                #RSSI over time at any distance
                else {

                        #Smoothed RSSI values from EWMA method
                        $ewma2 = ($α ∗ $rssiValue) + ((1 − $α) ∗ $ewma_prev);
                        #σ calculation using the variance formula
                        #and smoothed RSSI values

                        $variance_ewma = $α2 ∗ (($rssiValue−$mean_RSSI)^2 / $RSSI_values_number) + $δ
                        $δ = ((1 − $α2) ∗ $variance_ewma_prev);

                        $ewma_prev = $ewma2;
                        $smoothed_RSSI = $ewma2;
                                $variance_ewma_prev = $variance_ewma;
                        $σ = $variance_ewma;
                        $PL(d) = −($smoothed_RSSI);
                }

                $i = $i + 1;
}
        # The distance computation
        #The wavelength computation from the speed of the light and
        #frequency of the channel: 2.4 Ghz
        $λ = 3∗(10^8) / 2.48∗(10^9);
        #The reference path loss
        $PL(d0) = (($Pe∗$receiver_antenna_gain∗$emitter_antenna_gain∗($λ^2)) / (((4∗π)^2)∗($d0^2)));

        $resultat = ($PL(d)−$PL(d0)−$σ) / 10∗$η;
        # The corresponding distance
        $distance = (10^$resultat ∗ $d0);
```

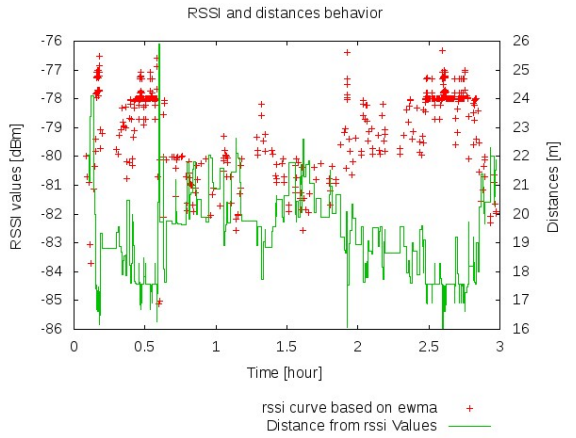
Where α and α_2 are the EWMA smoothing factor applying to calculate the smoothed RSSI values and variance. Using a constant smoothing factor (also named a weighted parameter) of value $\alpha = 0.01$ leads to an estimated distance whose value is closer to the real distance. This is valid for the three separation distances between the emitter sensor node and three receivers nodes. Moreover, we generally observed almost the same results when using $\alpha = 0.03$ (cf. Figure 3.37). However, it should be noted that the value of the used reference distance change the obtained results. Indeed, when applying the recommended reference distance of value $d_0 = 1$ m [22] an estimated too large distance is obtained compared to the current distance. In addition the same reference distance value can not be applied for a good estimate of the corresponding current distances of 3.65 m, 7.3 m and 18.69 m. For instance, with a current distance of 18.69 m, when applying $\alpha = 0.01$ or $\alpha = 0.03$, $Ge = Gr = 1$, $d_0 = 0.1$ m, we obtain an estimated distance of 18.3725 m. This resulted estimated distance value corresponds to the one that is



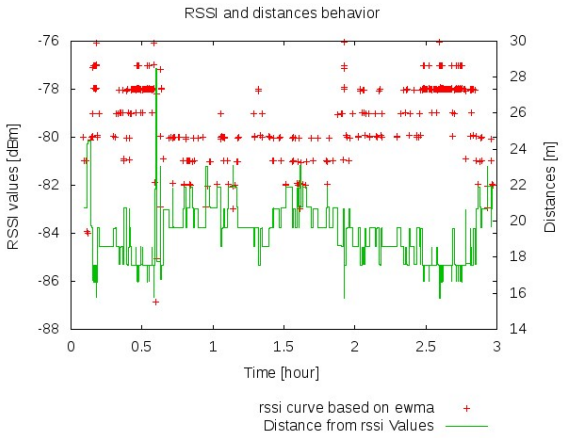
(a) Measured and non-filtered RSSI values at distance of 18.69 m



(b) Filtered RSSI values at distance of 18.69 m: $\alpha = 0.01$, $d_0 = 0.1$ m



(c) Filtered RSSI values at distance of 18.69 m: $\alpha = 0.03$, $d_0 = 0.1$ m

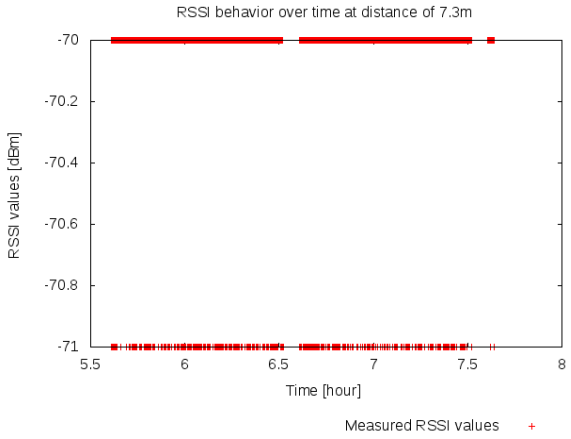


(d) Filtered RSSI values at distance of 18.69 m: $\alpha = 0.3$, $d_0 = 0.1$ m

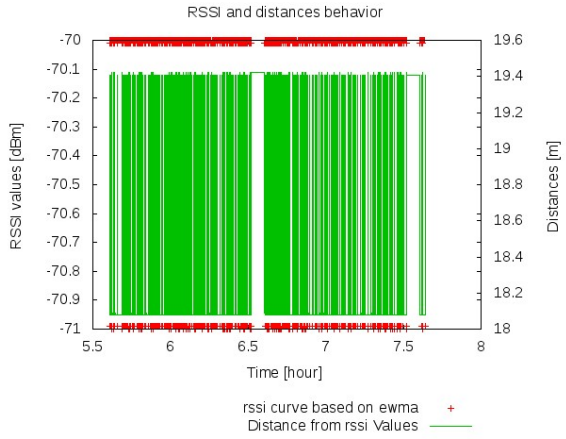
Figure 3.37: RSSI in function of the distance collected by a receiver node located in a different room and distant to 18.69 m from an emitter node

the most closer to the current distance when testing several values of α, d_0 and η. However, as shown in Figure 3.37 the same estimated distance value is obtained when applying a weighted parameter value of $\alpha = 0.3$. But this weighted parameter value of $\alpha = 0.3$ displays distances much closer to 18.69 m compared to $\alpha = 0.03$. Instead this result is not validated for the two other real distances of 3.65 m and 7.3 m while our goal is to find a weighted parameter value α that is adaptable to different indoor environments.

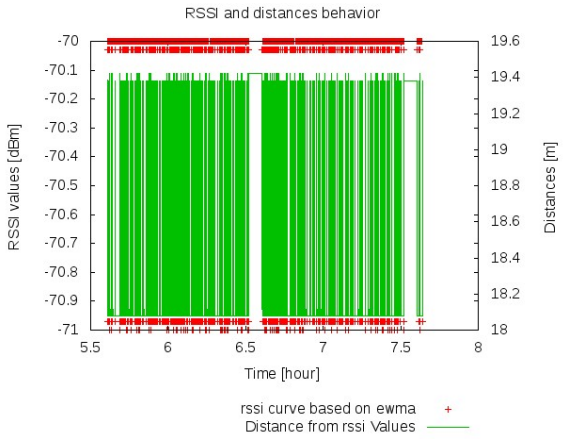
$\alpha = 0.03$ or $\alpha = 0.01$ and $d_0 = 0.04$ m lead to an estimated distance of 7.33 m which almost corresponds to the true distance of 7.3 m as reported in Figure 3.38. Instead, when applying a reference distance $d_0 = 0.1$ m as presented in Figure 3.38 we observe an erroneous too high estimated distance of value 18.35 m. We come up with this conclusion: the greater the reference distance decreases, better the result converges toward a better distance estimation. This hypothesis is also verified for the data measured at the real distance of 3.65 m as presented in Figure 3.39. Such figure shows an estimated distance of 4.9924 m. When $d_0 = 0.1$ m, we note an erroneously higher estimated distance of about 12.480 m as presented in Figure 3.39. A



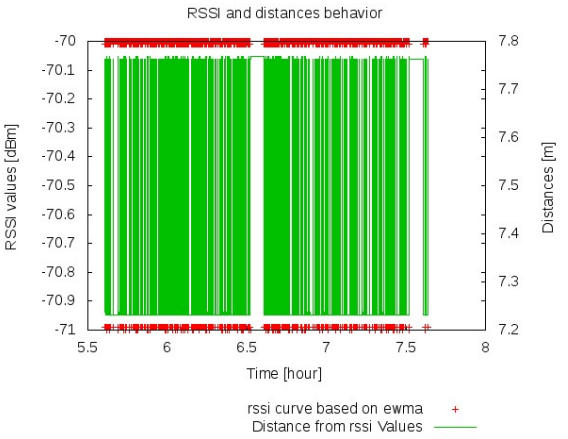
(a) Measured and non-filtered RSSI values at distance of 7.3 m



(b) Filtered RSSI values at distance of 7.3 m: $\alpha = 0.01$, $d_0 = 0.1$ m

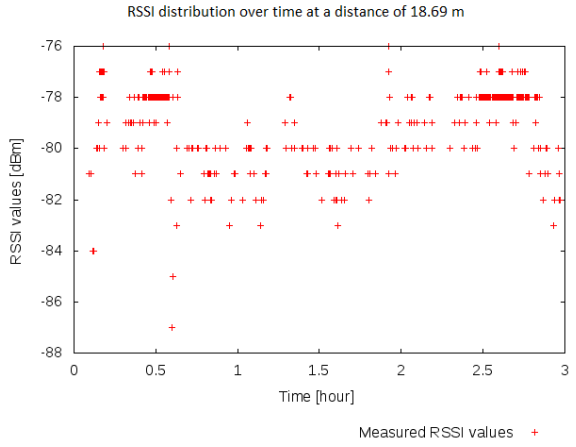


(c) Filtered RSSI values at distance of 7.3 m: $\alpha = 0.03$, $d_0 = 0.1$ m

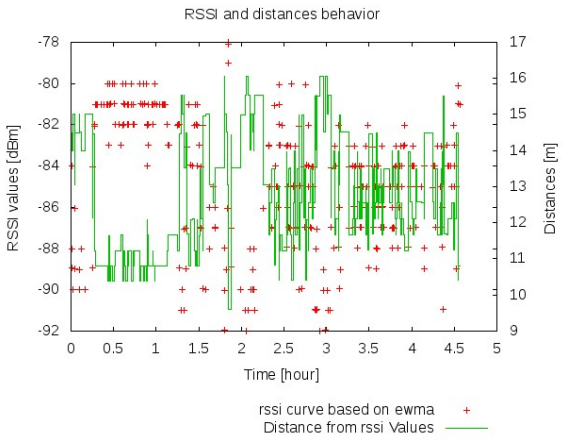


(d) Filtered RSSI values at distance of 7.3 m: $\alpha = 0.01$, $d_0 = 0.04$ m

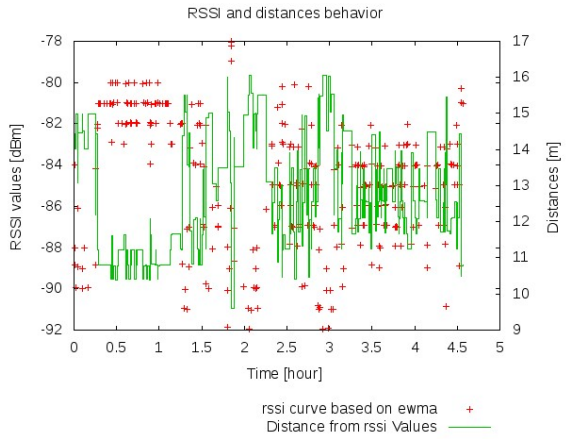
Figure 3.38: RSSI in function of the distance collected by a receiver node in the same room and distant to 7.3 m from an emitter node



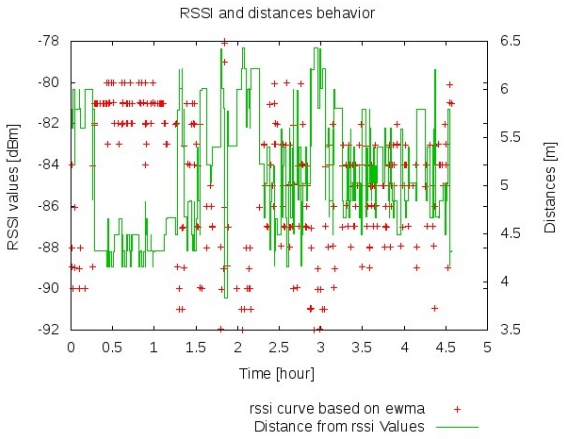
(a) Measured and non-filtered RSSI values at distance of 3.65 m



(b) Filtered RSSI values at distance of 3.65 m: $\alpha = 0.01$, $d_0 = 0.1$ m



(c) Filtered RSSI values at distance of 3.65 m: $\alpha = 0.03$, $d_0 = 0.1$ m



(d) Filtered RSSI values at distance of 3.65 m: $\alpha = 0.01$, $d_0 = 0.04$ m

Figure 3.39: RSSI in function of the distance collected by a receiver node in the same room and distant to 3.65 m from an emitter node

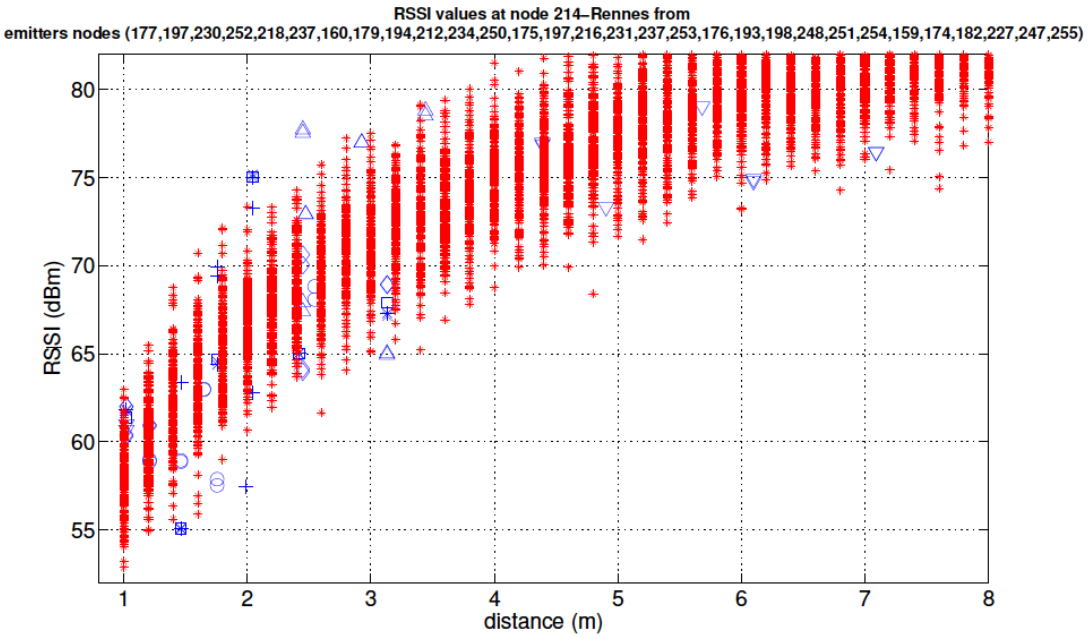


Figure 3.40: RSSI distribution at one receiver node on data coming from 30 emitters nodes at SensLAB Rennes platform testbed during 20 minutes: blue geometric patterns are measured RSSI and red asterisk pattern are a fitted log-normal distribution shape

better result requires a more smaller reference distance value. The obtained estimated distance of 4.9924 m is even an inappropriate estimation according to the real distance of 3.65 m.

With EWMA, we have demonstrated the need to take into account different values of reference distances according to three different positions of three receiver nodes in two different environments to estimate the corresponding distance from the measured RSSI. The reference distance value is taken into account when computing the value of the reference path loss. This shows at the same time the non-conformity of data smoothing methods based on exponential weighted moving average to smooth the measured RSSI values for adaptable technical filtering in any indoor environment. In fact, there are lots of obstructions which are introduced by shadowing effects while disrupting the direct path between a source named an emitter and a destination i.e a receiver. Obstacles can be obstructing objets in the office, furniture, people moving inside the environment. The direct path is therefore recovered with other paths called non line of sight paths. Recall from our experimental results that the constant smoothing factor value of $\alpha = 0.01$ or $\alpha = 0.03$ are the values of the constant smoothing factor that lead to better results compared to the other parameters in the interval $[0; 1]$. $\alpha = 0.01$ or $\alpha = 0.03$ better fit the original curve and result in less errors on distance estimation. The use of time series analysis helps to better characterize variations observed on measured RSSI values while confirming the dominant effect is the multipath.

Furthermore we realized several experiments in SensLAB Rennes platform testbed. Achieved results show that even if data are mostly altered due to obstructing factors in the indoor environment, there can be some distance information as the RSSI distribution almost follows a log-normal distribution as shown in Figure 3.40.

3.5 Summary and discussion

Based on deployment strategy, we have conducted several experiments in which we have studied in detail RSSI behavior in indoor environment. From experimental results, even with a quasi absence of antenna directivity and not entirely omnidirectional radiation pattern, irregularities and high dispersion obtained from reported RSSI values depend on both: the deployment orientation and position of the antennas of the emitter/landmark node and receiver/blind mobile node. This assumption is confirmed by experiment in an anechoic chamber and in several realistic indoor testbeds. In short, what is important is the relative orientation of the antennas one with regards to the other. This result remains valid in all the testbeds we collected data on and are considered for all kind of devices. Results also affirm the electronics devices we manipulate are not fully omnidirectional, nor errorless. Consequently, we suggest an explicit validation of datasheet technical information by performing additional tests in case of electronic sensor devices with antennas features. The proposed RSSI-based localization algorithms presented in Chapter 4 integrate these results. Recall, in a IEEE 802.15.4 network, a landmark is an emitter node with known position. A blind mobile node is a receiver node trying to determine its unknown position or coordinates.

We have experimented what should be the behavior of RSSI values in small time periods. Results in all testbeds report very dispersed and fluctuated RSSI readings. Different ranges of RSSI values provide distance information with different levels of accuracy. We also studied the difference between a large and a small room. Our experimental studies have shown a certain dependence on the material: RSSI readings are also affected by hardware quality and the battery pack of Tmote Sky node produces attenuation.

Other independent experiments conducted in a corridor and in three different testbeds size in indoor environments and presented in Section 3.3 demonstrate the necessity to learn each emitter node separately; if not, there is no correlation between measured RSSI values and distances. On the receiver node, only when considering the RSSI readings separately from each emitter node, a behavior that matches the log normal shadowing model pattern is able to be verified: RSSI decreases as a function of distance. Considering RSSI readings from each landmark separately, we were able to detect that, an emitter node among the others can be affected by abnormal behavior due to its environmental location. As RSSI is greatly affected by the environment, a node may not be in a good propagation conditions. In consequence, when considering all together all emitter nodes regarding the same receiver node position, RSSI readings on a receiver from data packets sent by this abnormal node could be a lot dispersed compared to the others. Consequently the overall distribution may be biased and thus deviates from normal pattern.

According to our experimental results and those of Heurteufeux *et al.* [7] conducted on SensLAB large scale experimental sensor nodes platforms, RSSI over a single link is not always bidirectional (about 10 %) even if communication links are mostly symmetrical when using 0 dBm as a transmitting power. Indeed, because of multipath, attenuation is different depending on how emitter and receiver nodes are oriented with respect or not to the furniture. This suggests and confirms a need to take into account in the localization algorithm the environmental orientation of the landmarks and of the receiving blind node antennas. We also demonstrated and argue the parameters of the propagation model should be estimated according to each emitter node environmental location and even more those values should be re-estimated for any environment and not kept constant.

In the presence of multipath, we additionally validated through measurements from different indoor testbeds 3.4.4 the non-conformity of the use of data smoothing methods based on expo-

nential weighted moving average to smooth the measured RSSI values for adaptable technical filtering in any indoor environment.

The analysis of corresponding results in different indoor testbeds emphasizes the behavior of RSSI values in indoor environments to not fully correlate with the use of the log-normal shadowing propagation model. In fact, the usage of the log normal shadowing model suggest the RSSI value to decrease with distance, on the other words attenuation is an increasing linear function of the distance. The log-normal shadowing propagation model is supposed to be more relevant then what we observe: there is somehow existing behavior that matches the log normal shadowing model pattern when separating the RSSI readings according to each emitter node position; RSSI decreases as a function of distance. This assumption is remarked by the whole set of results presented in this document. However, the dispersion between RSSI values is still very important and cannot lead to good accuracy on the prediction of the distance computed from the linear regression. Other contributions like [137, 54, 68] consider that, in an indoor context, signal attenuation increases monotonously with the distance. That observation may not be valid anymore in presence of multi-path propagation effects. We proposed to alleviate the multipath effects by implementing an alternate model named the Biased Log-Normal Shadowing Model or B-LNM.

Chapter 4

Proposed RSSI-based one hop localization algorithms: Indoor localization accuracy

Contents

4.1 Introduction

Measurement campaigns and results obtained from several indoor experimental testbeds demonstrate the greater variability of the RSSI measure because of the irregular pattern of the indoor radio channel propagation. Several parameters should be take into consideration when designing an indoor localization algorithm: the layout of the room, the smaller covered distance, the construction materials, the furniture pieces and their arrangement, the obstacles, the non-omnidirectional pattern antenna. The radiation pattern drawn in Chapter 3 is not entirely

uniform, a few directions experience a strong attenuation due to electronics on one of the board. These are multiple factors that impact the ranging error or the estimation of the separation distance between a wireless emitter node and a wireless receiver node.

Consequently some deployment positions are more favorable than others: one deployment position of the wireless sensor nodes antennas (*TowardCenter perpendicular*) have a better polarization compared to the others. That is the emitter antenna deployment with respect to a receiver antenna should additionally be considered, in terms of identifying the emitter position which provides a better linear relationship between measured RSSI values (*i.e.* less variation on reported RSSI values) and the corresponding separation distance.

Among the four studied deployment positions with respect to a wireless emitter node antenna and a wireless receiver node antenna, one deployment position (*TowardCenter perpendicular*) acquires a better polarization, less dispersion of RSSI values and best linearity between RSSI and distance. This deployment is used in our present localization architecture scenario. Despite, notice that as presented in Chapter 3, the other deployment strategies while showing few more fluctuations and less polarization, still have nearly the same trend as the selected one.

However, the high difference between corner landmarks and other landmarks indicates that reflections on the walls, from furniture pieces have a similar effect as having an even less circular antenna diagram. In fact, depending on the location in the room, the behavior of the signal propagation differs greatly. These results and the ones from [133, 147, 134, 135, 5] confirm that most actual wireless sensor node cannot be considered as omnidirectional in an indoor scenario.

Furthermore, our actual results in different indoor environments (cf. Chapter 3) have demonstrated the localization performance depends on the environment and the used model to infer distance from attenuation. Moreover, in addition of accumulated errors (ranging errors) during the distance measurement phase, the localization calculation techniques have an impact on the final localization accuracy.

Classical localization algorithms make use of a set of reference nodes, usually called *landmarks*, which know their own position and transmit this information to the mobile node.
A mobile willing to find its own position can then combine these coordinates with, for example, evaluation of the distance that separates it from each landmark and applies then trilateration or multilateration techniques. If landmarks are not aware of their own position, the computed position can still be deduced but it will remain relative to the local scene.

Recall, since no additional devices and no synchronization between nodes are needed compared to AOA and TOA approaches, the RSSI ranging measurement method involves an overall low-cost wireless sensor network localization approach.

However as demonstrated in Chapter 3, the range-based RSSI measurement in indoor, is mostly affected by environment variations and multipath fading. Acquiring ranging information hereafter, a localization computation technique uses such information as input to compute the corresponding 2D or 3D coordinates. As mentioned in Chapter 2.2, in a case where a node is in the communication range of another node, both nodes are neighbors each other and their distance separation can be estimated as one-hop ranging.

However, the achieved accuracy of most range-free algorithms are not well suitable for the use in indoor environment. Achieved accuracy is around more than 3 m using more than 80 landmarks for instance with DV-hop in [29]. Accuracy of 2 m is reported in [92] with 75 landmarks in a map of dimensions 50 m by 50 m.

Both algorithms use the radio connectivity to compute the number of hops as the localization

metric. In [86], ROCRSSI only compares raw measured RSSI values to find out the radio pro-ximity between the nodes. Evaluated in an indoor testbed of size 40 m by 30 m, the ROCRSSI algorithm reports an average localization error of 5.18 m using three landmarks and 11 receiver nodes positions. In addition, most of these range-free algorithms are obtained in simulation. Besides, as discussed in Chapter 3 and demonstrated in [127, 123, 136, 5], most of the simulated indoor RSSI-based localization algorithms achieve poor accuracy when deployed in real environment as the indoor environment is a dynamic and unpredictable environment in which the radio signal varies greatly.

Furthermore, in most proposed RSSI-based ranging algorithms using connectivity information [7], the accuracy is not good enough for implementation in indoor environments (e.g. a 2D localization error of about 2.2 m is obtained in 11 m by 9 m indoor area) when regarding the number of used landmarks.

Besides, because of indoor environment irregularities that affect the RSSI, most relevant proposed RSSI-based localization algorithms with ad-hoc or 802.15.4 wireless sensor architectures summarized in Table A.1 in appendix A do not provide a sufficient localization accuracy (cf. Table A.2).

Consequently the RSSI-based indoor localization accuracy still need to be improved. As a result in this chapter we consider the design of RSSI-based indoor localization algorithms. However, a design of such localization algorithm is related to the consideration of different constraints: the type of positioning mechanism (one-hop or multihop) and the amount of communication messages needed to perform localization within the wireless sensor network under the requirement of energy conservation. We should also consider the choice between a local or a global coordinate system, such as the global positioning system (*i.e.* GPS). However, recall that GPS signals are mostly attenuated in indoor environment. GPS receivers are acquired at an expensive cost and use high amounts of energy due to the GPS continuous location update [28, 31, 158]. Consequently, GPS is not suitable to be used in wireless sensor networks for indoor location. In the case of local coordinates systems, there can be a need to have some reference nodes with known positions.

We are looking for adaptable localization mechanism to compute one or more positions from the estimated distances based on RSSI measurements. This problem is considered from different points of view: geometrically by solving linear system of equations like trilateration and triangulation [159, 5, 6, 88], multilateration and multiangulation [88, 1, 20] and min-max [30, 160, 6]; statistical machine learning [123, 161, 162]; statistical point of view using signal processing with a parametric model such as Maximum likelihood [163, 5, 164, 6]; analysis of a vector subspace of matrix (e.g. analytical methods of principal component like Multi Dimensional Scaling [165, 166, 167, 1, 168, 169, 170, 125]); linear optimization methods such as semi definite programming (SDP) [168] and Discrete Stochastic Approximation (DSAL) [171, 33].

The common solution to all these mechanisms is to estimate one or several positions at the same time in a network using the measured RSSI values and a propagation method to map measured RSSI to corresponding distances. The estimated absolute position(s) is (are) generally relative to the position of one or several sensor nodes selected as reference and named landmarks or anchor nodes. In fact, the distance is a relative measure that does not depend on the considered reference system.

After presenting each mentioned localization mechanism, we first implement and evaluate a Maximum Likelihood Estimator (MLE)-based localization algorithm [5, 6, 164] to find out the accuracy that can be achieved in a realistic indoor testbed. These help us to quantify the

loca-lization accuracy. MLE is a signal processing localization mechanism based on a statistical model. MLE-based localization can compute one position at each execution with some landmarks.

We obtain a localization accuracy of 1 m in an indoor semi-furnished office testbed of size 4 m by 3 m. We then propose three other more accurate one-hop localization mechanisms/algorithms for low-power and low-cost wireless sensor nodes.

We also implement a fully distributed one-hop RSSI based localization algorithm in which we introduce the concept of virtual landmark. We define a virtual landmark as a blind node whose RSSI values are involved in the localization process of its neighbors blind nodes. In addition of RSSI values received from neighbors virtual landmarks, a blind node uses the measured RSSI values as well as the 2D coordinates from neighbors landmarks. Finally the blind node is able to compute itself its $2D$ coordinates and those of its neighbors. This distributed algorithm only needs at least two landmarks and some neighbors virtual landmarks to be implemented. However, in order to compare under the same conditions our proposed algorithm with multilateration and multi dimensional scaling, we use four landmarks which is the minimum number of landmarks required by localization with multilateration approach. The chapter is organized as follows: in Section 4.2 we present the state of the art of localization mechanism. We then validate each presenting localization algorithm/mechanism based on the theoretical log-normal shadowing model. Section 4.2.4 is devoted to the proposed one-hop localization mechanism. Each proposed one-hop localization mechanism is described, evaluated and compared with the aforementioned localization mechanisms in realistic local indoor testbeds. Simulation and experimental results (in remote SensLAB large scale wireless sensor network) from distributed localization is presented in Section 5.2. Discussion and summarization of the experimental results are progressively provided.

4.2 Reference localization algorithms

Many different techniques have been proposed to derive a position from RSSI measurements. The RSSI measurements are commonly modeled as log-normal random variables (cf. Chapter 3 of this thesis document). Before computing the unknown position of a blind mobile node within a wireless sensor network, we first need to select a propagation model in order to map a distance from measured RSSI values. RSSI is viewed as the attenuation experienced by radio packets measured by sensor nodes depending on the distance. This propagation model should represent a good compromise between accuracy and complexity. Indeed, the algorithm has to be performed on embedded devices and hence should not require too much computation or too much data exchange. As mentioned in Chapter 3, among the presented propagation models, the Log-Normal Shadowing Model (LNSM) [24], often used to model indoor propagation represents a good compromise between complexity and feasibility. Recall that this model relates the path loss between an emitter and a receiver $\mathrm{PL}(d)$ to the distance that separates these two nodes (d) according to the following equation (4.1):

$$\mathrm{PL}(d) = P_e + \mathrm{PL}(d_0) + 10.\eta.\log_{10}\left(\frac{d}{d_0}\right) + \mathcal{N}(0,\sigma^2)\,, \tag{4.1}$$

where P_e is the emitted/transmitted power, $\mathrm{PL}(d_0)$ is a reference path loss (in dBm) measured at a distance d_0 (usually 1 m), η is the attenuation factor of the environment, and $\mathcal{N}(0,\sigma^2)$ is a gaussian random variable with zero mean value and standard deviation σ. $\mathrm{PL}(d_0)$ is a constant

calibration offset, η reflects the average attenuation of the environment and σ^2 reflects the effect of shadowing (slow fading).

Recall again that the RSSI measured by a wireless receiver is the difference between the transmission power and the corresponding path loss. In most of our experiments, as the transmission power is fixed to $0\,dBm$, the RSSI is simply the opposite of the path loss: $RSSI(d) = -\mathrm{PL}(d)$.

Under this definition, when a wireless sensor node measures a RSSI value of R, the random variable r which defines the variable $RSSI(d)$ from the model in Equation (4.1), can be expressed with a Gaussian distribution as $r \sim \mathcal{N}(\,-\mathrm{PL}(d_0) - \eta 10\log_{10} d\,,\,\sigma^2)$ whose probability density function is:

$$f(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(r+\mathrm{PL}(d_0)+10\eta\log_{10} d)^2} \tag{4.2}$$

Thus, this parametrical model is used to derive the following parameters: the path loss η of the environment, the reference power loss $\mathrm{PL}(d_0)$ and the variance σ^2) of the random effect of shadowing. The shadowing effect could be due to the presence of an obstacle in the middle of the communication link disturbing at the same time the direct path between an emitter node and a receiver node. However, from the results of our experimental testbeds, if the RSSI values of the emitters are not learned separately on the receiver, we would not verify the hypothesis that the RSSI values decrease as a function of the distance within log-normal shadowing propagation model in indoor environment. Thus, the log-normal shadowing propagation model parameters are learnt separately from each emitter node position being in the communication range. In addition, for the distributed algorithm, those parameters are learnt separately depending on a blind node position. During the localization process, such parameters should be computed before to derive the distance that separates any emitter node to any receiver node. Moreover, in our strategy this pre-processing step is included in the localization process without a need of a calibrating phase as did in [54, 172, 3].

4.2.1 Propagation model parameters estimation

The values of the log-normal shadowing propagation model parameters ($\mathrm{PL}(d_0)$, η, σ^2) are estimated using a set of RSSI measurements coming from the landmarks. From our localization algorithm architecture point of view, the blind mobile node is responsible at each unknown location to collect position-related information which includes RSSI values, landmarks identifiers and 2D coordinates, the sequence number of the received packet, etc.

While receiving position-related data, a blind node can compute online the propagation model parameters of the corresponding environment using the regression model equations (4.5) and (4.7) based on maximum likelihood estimation. The estimators for $\mathrm{PL}(d_0)$ and η that correspond to the regression parameter estimation is based on the least square criterion. The method of ma-ximum likelihood is chosen to estimate those values that are more consistent with the sample data [173]. After estimating the propagation model parameters, the blind node then estimates its $2D$ coordinates using the landmarks $2D$ coordinates and the RSSI readings from such landmarks. From the algorithm point of view, the propagation model parameters estimation step is not an offline phase that is separated from the position estimation phase as realized in many works [99, 3].

In addition, as demonstrated in our experimental results during measurement campaigns in Chapter 3, such parameters depend on the environment and where the landmarks are placed

with respect to the blind node. However, in previous research works a fixed path loss exponent is used in the radio propagation model whatever the environment is. Only in the recent works of LU *et al.* [99], the path loss exponent η is considered to vary with the environment. The authors have demonstrated that the localization error (*i.e.* RMSE) is reduced from 4. 45 m when a fixed path loss exponent is chosen to 2.13 m when using their method. The authors assumed an indoor environment of size 18 m by 18 m using their Polygon method.

We have proposed and have implemented online localization mechanisms that first compute the propagation model parameters according to each landmark environmental location using a set of measurements coming from a subset of known positions that we call the *learning data set*. Then, these estimated parameters are considered to find out the corresponding unknown positions. These two-online steps included in any blind mobile node positioning process is different from calibration. In fact these steps are computed online during the localization process and not separated from the localization phase.

In most cases when using a calibration, RSSI values are required to be collected during an offline phase using several sets of measurement points in the corresponding environment. These values are then used to compute the propagation model parameters being the same constant value (one estimated value for $\mathrm{PL}(d_0)$, one for η and one for σ) for any involved landmarks. In our approach, regarding one-hop localization algorithms, propagation model parameters are first estimated by the blind mobile node itself according to each landmark position. Then, such parameters are used by the same blind mobile node to estimate its unknown position within the environment. Note that the number of positions and RSSI data to use for propagation model parameters estimation are different to those used by the blind mobile node to estimate its own position.

We assume the log-normal shadowing model (cf. the equation 4.1) for T RSSI values measured by a wireless sensor node from signals coming from N different sensor nodes whose corresponding distances $\{d_1, \ldots, d_N\}$ are known. The goal is to compute the set of parameters $\hat{\mathrm{PL}}(d_0)$, $\hat{\eta}$ and $\hat{\sigma}^2$ that characterize the environment. Let us define the T RSSI measurements as a vector $\boldsymbol{r}_n = (r_n(1), \ldots, r_n(T))$ associated to each distance d_n. In this case, the random vectors $\boldsymbol{r}_n$ are considered to be independent and identically distributed (*i.i.d*) according to a Gaussian law of dimension T whose components are also considered independent of each other.

If each component $r_n(t)$ of the vector $\boldsymbol{r}_n$ is considered as a random variable with distribution $\mathcal{N}(\,-\mathrm{PL}(d_0) - \eta 10\log_{10} d_n\,,\,\sigma^2)$, with probability density function as in equation 4.2, the probability density function of the vector $\boldsymbol{r}_n$ is expressed as:

$$f(\boldsymbol{r}_n) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(r_n(t)+\mathrm{PL}(d_0)+10\eta\log_{10} d_n)^2} = \frac{1}{(2\pi\sigma^2)^{T/2}} e^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(r_n(t)+\mathrm{PL}(d_0)+10\eta\log_{10} d_n)^2} \tag{4.3}$$

From the set of the N random vectors $\boldsymbol{r} = \{\boldsymbol{r}_n\}_{n=1,\ldots,N}$ and when each of them uses the definition in equation (4.3) above, the joint probability density function of these N random vectors is obtained as:

$$f(\boldsymbol{r}) = \prod_{n=1}^{N} f(\boldsymbol{r}_n) = \frac{1}{(2\pi\sigma^2)^{NT/2}} e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{t=1}^{T}(r_n(t)+\mathrm{PL}(d_0)+10\eta\log_{10} d_n)^2} \tag{4.4}$$

If a sensor node collects the NT RSSI values as a set of N vectors

$\{(R_1(1),\ldots,R_1(T)),\ldots,(R_N(1),\ldots,R_N(T))\}$ and when assuming the parametric model defined in equation (4.4), we estimated the set of parameters that maximize the likelihood between the measured values and the values defined by the model. Applying the logarithm to the expression (4.4), the problem to solve is finally expressed as:

$$(\hat{\mathrm{PL}}(d_0),\hat{\eta},\hat{\sigma}^2) = \max_{(\mathrm{PL}(d_0),\eta,\sigma^2)} -\frac{NT}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(R_n(t)+\mathrm{PL}(d_0)+10\eta\log_{10}d_n\right)^2$$

Rewriting the above expression permits to estimate the propagation model parameters ($\{\mathrm{PL}(d_0),\eta\}$ and σ^2) separately. The estimation problem is summarized under a vector form following the vector expression below:

$$(\hat{\boldsymbol{\alpha}},\hat{\sigma}^2) = \max_{(\boldsymbol{\alpha},\sigma^2)} -\frac{NT}{2}\ln(2\pi\sigma^2) - \frac{T}{2\sigma^2}||\boldsymbol{R}-\boldsymbol{d}\boldsymbol{\alpha}||^2$$

where :

$$\boldsymbol{R} = \begin{pmatrix} 1/T\sum_{t=1}^{T}R_1(t) \\ \vdots \\ 1/T\sum_{t=1}^{T}R_N(t) \end{pmatrix} \quad \boldsymbol{d} = \begin{pmatrix} 1 & 10\log_{10}d_1 \\ \vdots & \vdots \\ 1 & 10\log_{10}d_N \end{pmatrix} \quad \boldsymbol{\alpha} = \begin{pmatrix} \mathrm{PL}(d_0) \\ \eta \end{pmatrix} \tag{4.5}$$

From equation (4.5) and denoting $\boldsymbol{d}^t$ the transpose of vector $\boldsymbol{d}$, we derive the log-normal shadowing propagation model parameters using the maximum likelihood estimator which also correspond to the regression parameters estimation based on least squares:

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{d}^t\boldsymbol{d})^{-1}\boldsymbol{d}^t\boldsymbol{R} \quad \text{then :} \quad \hat{\mathrm{PL}}(d_0) = \hat{\alpha}(1),\ \hat{\eta} = \hat{\alpha}(2) \tag{4.6}$$

$$\hat{\sigma}^2 = \frac{1}{TN}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(R_n(t)+\hat{\mathrm{PL}}(d_0)+\hat{\eta}10\log_{10}d_n\right)^2 \tag{4.7}$$

This simplified estimation of σ^2 can be extended to the case where we consider different additive noise levels (that may disturb) for each link between sensor nodes. In such case, there would be N different values of estimated σ^2 as:

$$\hat{\sigma}^2 = \begin{pmatrix} \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_N^2 \end{pmatrix} = \begin{pmatrix} 1/T\sum_{t=1}^{T}\left(R_1(t)+\hat{\mathrm{PL}}(d_0)+\hat{\eta}10\log_{10}d_1\right)^2 \\ \vdots \\ 1/T\sum_{t=1}^{T}\left(R_N(t)+\hat{\mathrm{PL}}(d_0)+\hat{\eta}10\log_{10}d_N\right)^2 \end{pmatrix} \tag{4.8}$$

with each entry of the vector $\boldsymbol{\sigma}^2$ being σ_n^2 defined as : $\hat{\sigma}_n^2 = 1/T\sum_{t=1}^{T}\left(R_n(t)+\hat{\mathrm{PL}}(d_0)+\hat{\eta}10\log_{10}d_n\right)^2$ for $n = 1,...,N$. We can then obtain the

maximum likelihood estimator (MLE) of any distance associated to measured RSSI from signals emitted by any landmark node.

We consider the log-normal shadowing model parameters and we obtain the estimated values by maximum likelihood (cf. equation 4.7) at each sensor node from a network of 16 selected nodes in SensLAB[1] large scale wireless sensor platform (Rennes platform). These values are summarized in Table 4.1. The average estimated values in the network composed of $N = 16$ wireless sensor nodes whose configuration is given in Figure 4.4) are $\mathrm{PL}(d_0) = -61$ dBm and $\eta = 2.1$. Note that, such values also correspond to the typical indoor environment values [25].

Table 4.1: Estimation of the propagation model parameters by maximum likelihood estimator (cf. the equation 4.5 and 4.7) using the RSSI values from an experiment when 16 SensLAB Rennes sensor nodes are selected as shown in Figure 4.5.

ID	214	176	178	193	198	218	225	232	247	249	252	253	157
$\boldsymbol{PL(d_0)}$	-52.95	-60.05	-50.18	-59.88	-61.94	-59.60	-66.75	-56.31	-67.53	-52.69	-66.42	-63.03	-62.74
$\boldsymbol{\eta}$	3.54	2.02	4.22	2.04	1.59	2.01	1.29	2.80	0.66	3.21	0.72	1.74	1.66
$\boldsymbol{\sigma^2}$	17.55	8.55	20.19	22.18	12.36	9.16	10.10	13.02	14.09	31.36	12.43	9.81	15.67
ID	163	236	244										
$\boldsymbol{PL(d_0)}$	-66.27	-59.92	-58.95										
$\boldsymbol{\eta}$	0.94	2.06	3.14										
$\boldsymbol{\sigma^2}$	21.69	15.38	28.29										

In the following, we analyze the properties of the regression estimators when applied to the distance estimation process. Similarly to equation 4.7, we can deduce the maximum likelihood regression estimator (MLE) of any distance d associated to RSSI measurements from data packets received by a wireless sensor node. Assume that a wireless sensor node (receiver) measures T RSSI values, referred to the vector $\boldsymbol{R} = (R(1), \ldots, R(T))$, from another remote emitting wireless sensor node known as landmark and situated at distance d from the receiver wireless sensor node. Using equation (4.7), we additionally assume the parameters of the propagation model $\{\mathrm{PL}(d_0), \eta, \sigma^2\}$ being already estimated. We defined by $\hat{d}_b$ the biased maximum likelihood estimator of the distance calculated as:

$$\hat{d}_b = \max_d \; -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(R(t) + \mathrm{PL}(d_0) + 10\eta\log_{10} d\right)^2$$

$$\text{thus :} \quad \hat{d}_b = 10^{\frac{-\overline{\boldsymbol{R}} - \mathrm{PL}(d_0)}{10\eta}} \quad \text{with } \overline{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} R(t)$$

We evaluate the impact of the parameters of the maximum likelihood estimator on the estimated distance from its corresponding bias and variance values. Under the assumption that the parameters $\{\mathrm{PL}(d_0), \eta, \sigma^2\}$ are estimated, the maximum likelihood estimator of the distance $\hat{d}_b$ defined above is observed as a noisy version (multiplicative noise) of the true distance d. If we substitute the values of $\bar{\boldsymbol{R}}$ in the above expression by the values from the propagation model (cf. equation 4.1), we obtain the expression of the following distance estimator :

$$\hat{d}_b = d\, 10^{\frac{\xi}{10\eta}} \quad , \xi \sim \mathcal{N}(0, \overline{\sigma}^2) \quad \text{with : } \overline{\sigma}^2 = \frac{\sigma^2}{T} \tag{4.9}$$

The mathematical expectation $E[\hat{d}_b]$ of the corresponding biased estimator (cf. the equation 4.9) is computed using the property $a^x = e^{\ln ax}$ and $E[10^{\alpha\xi}] = 10^{\alpha^2\frac{\overline{\sigma}^2}{2}\ln 10}$ for a given constant α :

[1] `www.senslab.com`

$$E[\hat{d}_b] = Kd \quad \text{where :} \quad K = 10^{\frac{\sigma^2 \ln 10}{2T(10\eta)^2}} \tag{4.10}$$

An unbiased estimate $\hat{d}$ can be derived if a multiplicative term K^{-1} is added to the biased distance estimator (cf. equation 4.9) as:

$$\hat{d} = K^{-1}\hat{d}_b = K^{-1}\, 10^{\frac{-\overline{R}-\mathrm{PL}(d_0)}{10\eta}} = d\, 10^{\frac{\xi}{10\eta}}\, 10^{-\frac{\sigma^2 \ln 10}{2T(10\eta)^2}} \quad \Longrightarrow \quad E[\hat{d}] = d \tag{4.11}$$

The variances of both estimators, $var\{\hat{d}_b\}$ and $var\{\hat{d}\}$ (biased and unbiased respectively) are:

$$\begin{aligned} var\{\hat{d}_b\} &= E[\hat{d}_b^2] - E[\hat{d}_b]^2 = d^2K^2\left(K^2-1\right) \\ var\{\hat{d}\} &= E[\hat{d}^2] - d^2 = d^2\left(K^2-1\right) \end{aligned} \tag{4.12}$$

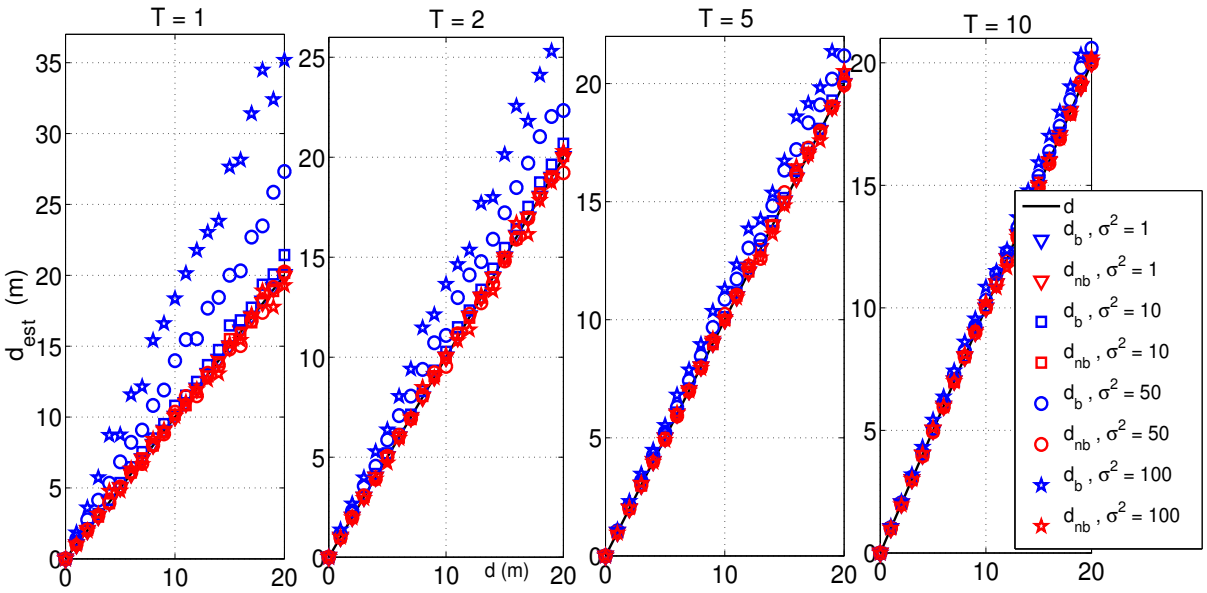


Figure 4.1: Expectations of biased (d_b in blue color) and unbiased (d_{nb} in red color) estimators of distances for the following additive noise values: $\sigma^2 = (1, 5, 10, 50, 100)$ and different values of the number of RSSI measurements $T = 1, 2, 5, 10$).

To illustrate the behavior of the two estimators, we first generate RSSI data under log normal model in Matlab to simulate a radio communication (RSSI measurements) between two nodes according to several distances from 1 m to 20 m with a step of 1 m. We draw a log-normal distribution of RSSI values from a 1000 Monte Carlo independent runs based on distances that separate two wireless sensor nodes.

Let us denote by $\hat{d}_b$ and $\hat{d}$, the respective biased and unbiased estimators of the distances d estimated with the maximum likelihood estimator. To analyze the effects of the values taken by both σ^2 and T (the number of the RSSI measurements from an emitter sensor node) on the estimated distances, we evaluate both the bias (Figure 4.1) and the variance (Figure 4.2) of

estimators $\hat{d}_b$ and $\hat{d}$ at the distances ranging from $[0, 20]$ m. More precisely, we study the impact of dealing with more or less noise and more or less number of RSSI measurements on the bias and the variance of the estimators of the distances.

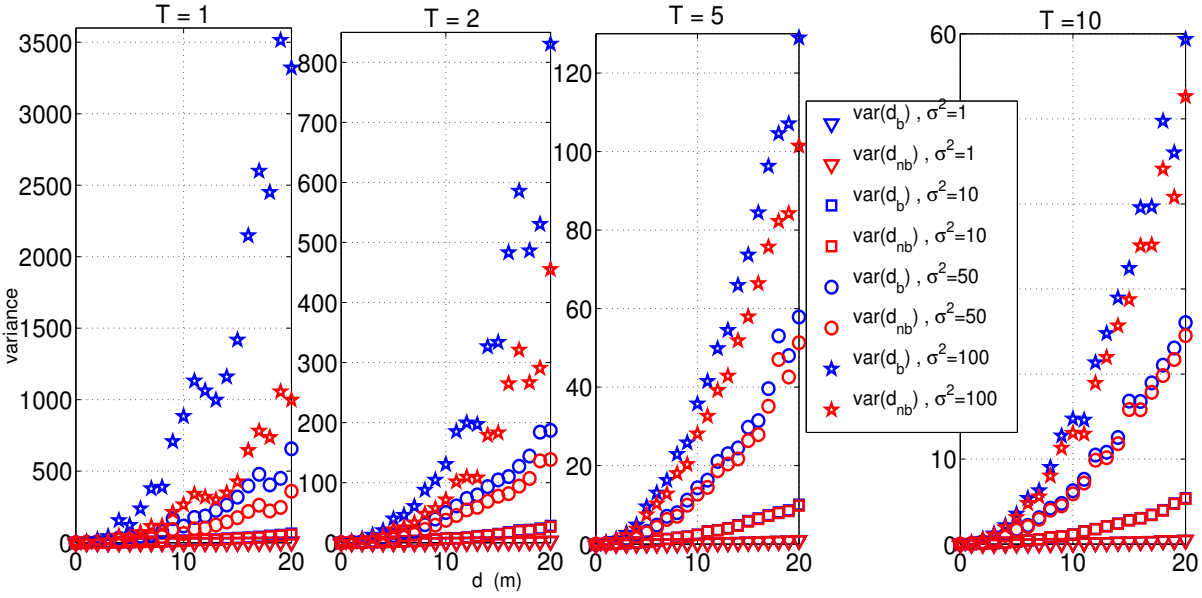


Figure 4.2: Biased (d_b in blue color) and unbiased (d_{nb} in red color) variance estimators of distances for the following noise values $\sigma^2 = (1, 5, 10, 50, 100)$ and different values of the number of RSSI measurements $T = 1, 2, 5, 10$).

We observe that, when considering for instance the unbiased maximum likelihood estimator, the distances are accurately estimated and close to the real distances values for any tested noise level ranging from $\sigma^2 = (1, 5, 10, 50, 100)$ and even with a single T RSSI measurement. The bias for the biased estimator $\hat{d}_b$ increases with σ^2 but it has a good performance when considering a minimum number of RSSI measures (for instance from $T = 5$). This can be explained by the fact that averaging decreases the gap around the real distance value. Such averaging (of RSSI measures) is considered under high levels of noise. In fact, in the expression of the variances (cf. the equation 4.12), K tends to 1 if the factor $2T(10\eta)^2$ increases more strongly than the value of σ^2. The variances of the two estimators have the same shape for all values of T and σ^2 to within a constant K^2 defined in equation 4.12. However, in general the variance of $\hat{d}$ is always smaller than the one of $\hat{d}_b$. But, the difference is still imperceptible when the values of the noise level σ^2 is below 10 and quite negligible for high values of σ^2 when considering $T = 10$ or more RSSI measurements.

Finally we illustrated in Figure 4.3 the law of the estimated distance error when considering both estimators: the biased $\hat{d}_b$ and unbiased $\hat{d}$ one. We choose for instance a real distance of $d = 1$ m and try to estimate it using both estimators. Note that the histogram resembles to a log-normal distribution as expected with a Gaussian shape for lower noise levels ($\sigma^2 = 1$) and when considering both number of RSSI measures ($T = 1$ and $T = 10$).

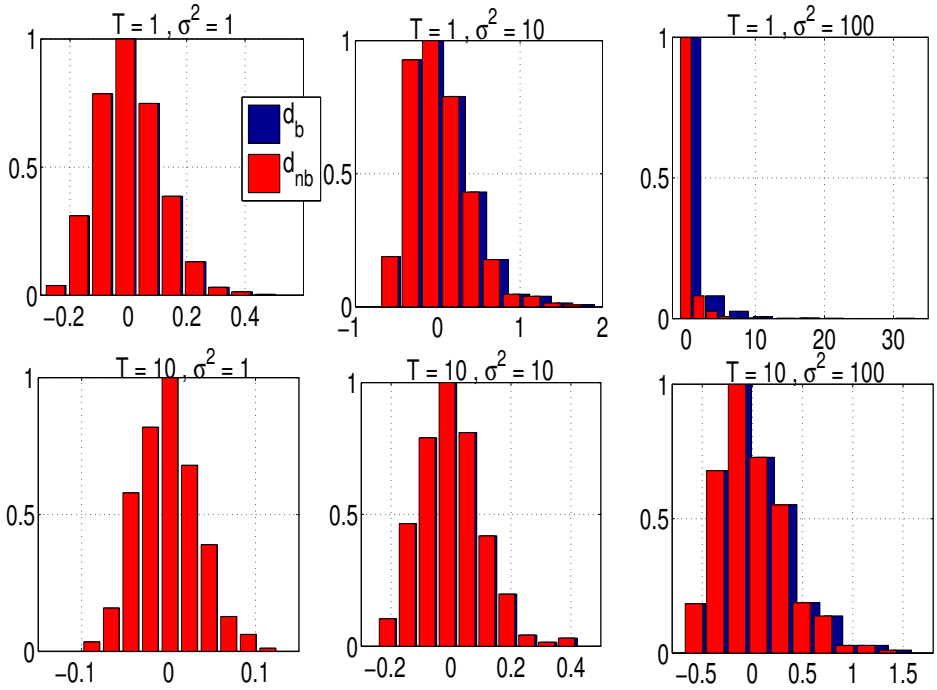


Figure 4.3: Histograms of the errors of both the biased $\hat{d}_b - d$ and unbiased estimators $\hat{d}_{nb} - d$ for different noise levels σ^2 (1, 10 and 100) ($d = 1$m) when first considering one RSSI measure $T = 1$, then several $T = 10$ RSSI measures to average.

From the histograms in Figure 4.3, we observe the distance estimator to behave like a noisy version of the true distance d:

$$\hat{d} = d + \zeta \quad \text{whose :} \quad \zeta \sim \mathcal{L}og\mathcal{N}(\mu\,,\, d^2(K^2 - 1))$$

In the case of the unbiased estimation, the noise is centered as ($\mu = 0$), while for the biased case the noise has a mean equal to the bias of the estimator ($\mu = E[\hat{d}_b] - d$).

Note that with the studied localization mechanisms hereafter, we will need to use the square of the distance. In this case, we consider the estimated values of the squared distances d^2 defined by $\hat{d}_{square}$. If the following expression in (4.11) is used as an estimator, it is biased by the term calculated with the equation (4.12) since the term $E[\hat{d}^2]$ is not equal to d^2 but $K^2 d^2$. We multiply each estimated distance $\hat{d}$ by the term K^{-2} to obtain an unbiased estimate of the square of the distance $\hat{d}_{square}$ as:

$$\hat{d}_{square} = 10^{\frac{-\overline{\boldsymbol{R}} - \mathrm{PL}(d0)}{5\eta}} K^{-2} = d^2\, 10^{\frac{2\xi}{10\eta}}\, 10^{-\frac{\sigma^2 \ln 10}{T(10\eta)^2}} \quad \Longrightarrow \quad E[\hat{d}_{square}] = d^2 \tag{4.13}$$

Using the property $a^x = e^{\ln a x}$ and $E[10^{\alpha\xi}] = 10^{\alpha^2 \frac{\sigma^2}{2} \ln 10}$ for a given constant α, the variance of the resulting estimator is:

$$var\{\hat{d}_{square}\} = E[\hat{d}^2_{square}] - d^4 = d^4 \left(K^8 - 1\right) \tag{4.14}$$

We have validated the theoretical model of the maximum likelihood estimator based regression parameter estimation by studying the impact of the acquisition parameters to the corresponding distance estimators.

We have noticed that even with additive noise on the RSSI values generated following a normal law (log-normal distribution), the unbiased maximum likelihood estimator is able to accurately estimate the distance. In the remainder of this thesis memory, in addition of other localization methods like trilateration and proposed methods, we will study the localization-based maximum likelihood estimator when firstly considering actual RSSI measurements on our local testbeds and secondly on SensLAB Rennes platform.

Localization accuracy is obtained using both empirical measures and 1000 independent Monte Carlo realizations (RSSI data generated with Matlab) for each localization algorithm. With Monte Carlo realizations, data are generated in Matlab according to a log normal distribution. Empirical RSSI measures are obtained in a network of 16 wireless sensor nodes whose configuration is shown in Figure 4.4. This wireless sensor node experiment is realized in SensLAB Rennes platform. Note that each sensor is associated to an identifier *NodeID* by the SensLAB Rennes platform. Such number are used during the post-processing phase.

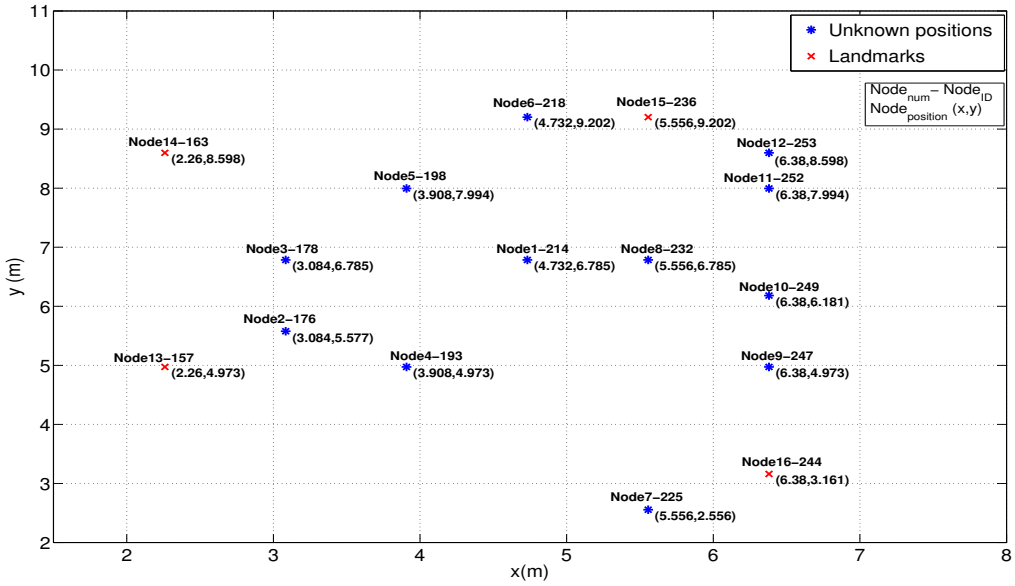


Figure 4.4: Architecture of the used wireless sensor nodes (total number of 16) in the SensLAB Rennes platform

In order to illustrate/validate theoretical model and empirical values, we will use for the remaining sections the unbiased distance and square distance estimators defined respectively by equations (4.11) and (4.13). We computed the first two moments (mean and variance) using data from one selected sensor node in the network in Figure 4.4. While selecting the first sensor (the node with identifier 214) we have computed the empirical bias and variances of the set of estimated square distances $\{\hat{d}_{square}(1,i)\}_{i=2,\ldots,16}$ as:

$$\begin{aligned}\text{Bias} =&(-0.002, -0.001, -0.0013, -0.002, -0.0083, -0.061, 0.0002, 0.0075, -0.003,\\ &0.009, -0.0077, 0.0085, 0.01, 0.014, 0.021)\\ \text{Variance} =&(0.62, 0.26, 0.55, 0.17, 1.32, 14.57, 0.016, 1.14, 0.34, 0.52, 1.30, 2.87, 2.80, 1.27, 7.91)\end{aligned}$$

The bias are almost null (near zero) as expected and their variances are near to the theoretical expression of the equation (4.14) whose real values are:

$$\{d_{1,i}^4 \left(K^8 - 1\right)\}_{i=2,\ldots,16} = (0.77, 0.32, 0.69, 0.20, 1.50, 15.14, 0.02, 1.58, 0.42, 0.77, 1.58, 3.88, 3.88, 1.87, 11.0386)$$

We run experiments remotely in the SensLAB Rennes platform through the website by selecting the list of sensors as shown in Figure 4.5 which corresponds to the configuration of Figure 4.4. Firmware are then uploaded on the selected nodes.

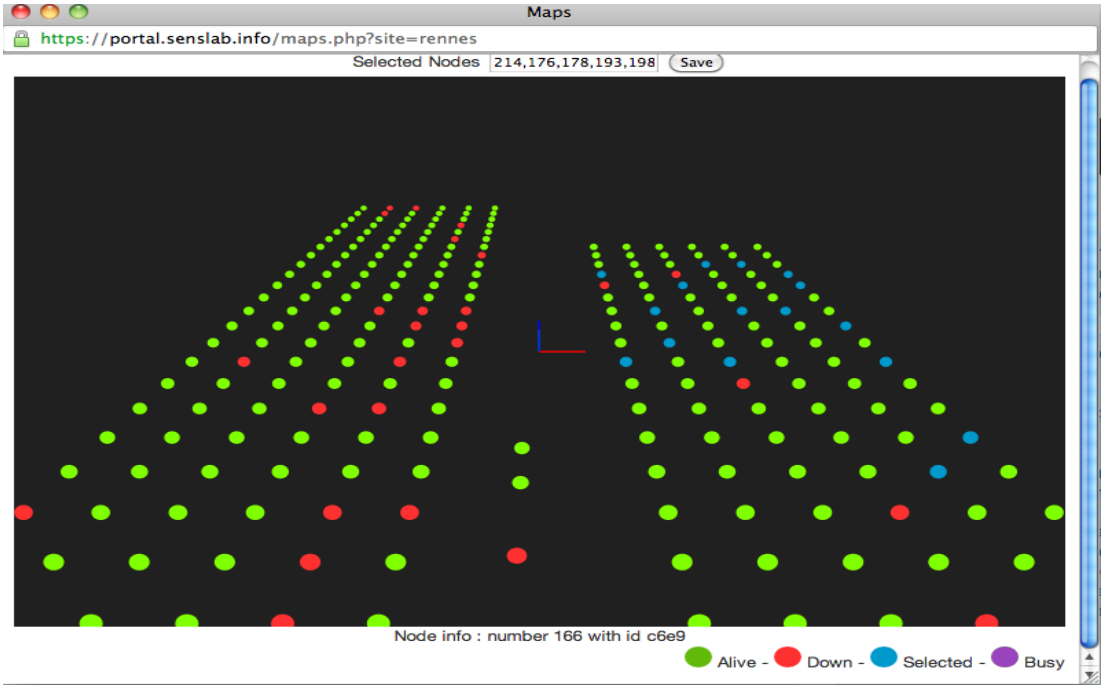


Figure 4.5: Network with 16 selected wireless sensor nodes (blue circles) deployed in SensLAB Rennes platform.

To illustrate whether there is a log normal shape, Figure 4.6 presents the RSSI values measured by 8 of the 16 sensor nodes (214, 176, 178, 225, 232, 163, 236 and 244 (cf. Figure 4.4)) depending on the real distance between each of the 8 nodes and the nodes related to the RSSI received packets during the acquisition time. We draw as a function of distances, the RSSI values measured between 8 sensor nodes knowing that each of the node has an emitter and a receiver feature. In most cases, the measured RSSI values do not decrease under a normal distribution as the distance increases.

Figure 4.7 details the number of packets received between 16 sensor nodes named: 157, 163, 176, 178, 193, 198, 214, 218, 225, 232, 236, 244, 247, 249, 252 and 253. We can observe that the sensor with the identifier 157 is not communicating (not receiving messages) with the sensor nodes 176, 218 and 252. No packet is received from the sensor 176 while this last is receiving packets from the other 11 sensor nodes. This can be explained by the fact that the sensor node 176 is not in a favorable propagation condition, thus the data message it sends is mostly affected by noise. Instead, the following sensor nodes with identifier 178, 193, 198, 225, 232, 247, 249, 253, 163 and 244 receive packets from all other sensor nodes except the one with the identifier 176.

In order to illustrate the latter distance estimator applied with real data, Figure 4.8 shows the

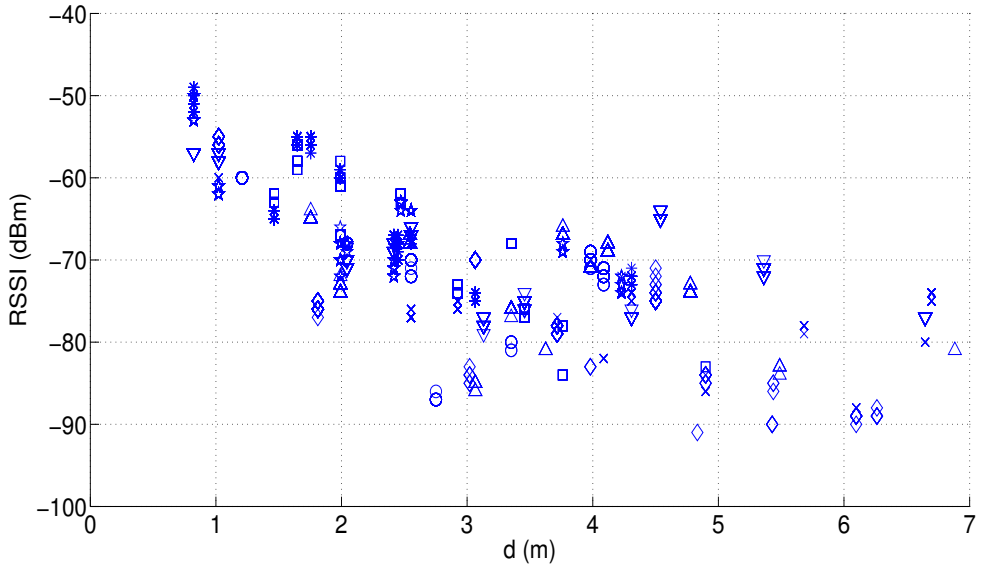


Figure 4.6: RSSI values measured from 600 data packets received at each sensor node depending on the distance between the sensor nodes. Each of the 8 different markers corresponds to a selected sensor node in SensLAB.

NodeID	214	176	178	193	198	218	225	232	247	249	252	253	157	163	236	244
214	0	0	616	563	514	493	617	501	543	546	87	539	476	0	487	18
176	1209	0	1299	0	1090	1118	22	22	55	13	1146	0	0	3	0	23
178	2107	0	0	500	492	412	42	434	32	442	216	111	540	211	444	17
193	1782	0	431	0	436	20	476	372	491	395	213	163	429	194	378	220
198	1779	0	476	439	0	415	449	374	391	31	229	425	392	214	380	6
218	1942	0	452	392	468	0	527	401	30	417	241	486	0	217	418	9
225	2756	0	42	623	73	598	0	564	645	245	20	39	46	14	37	298
232	1711	0	456	421	394	402	441	0	372	397	29	411	388	201	364	13
247	2257	0	38	533	481	32	622	464	0	509	225	533	20	23	34	229
249	1812	0	438	434	32	344	445	378	470	0	226	427	406	200	374	14
252	227	0	729	688	657	684	72	77	665	616	0	705	0	266	614	0
253	1848	0	254	441	443	408	45	386	472	413	251	0	403	221	398	17
157	2178	0	553	526	505	0	53	462	40	463	0	500	0	22	471	227
163	216	0	726	703	683	669	42	641	46	607	268	700	51	0	631	17
236	1652	0	416	411	414	392	405	352	332	357	251	422	402	194	0	0
244	328	0	79	1088	52	81	1329	667	1085	56	3	68	1078	17	69	0

Figure 4.7: Table of number of packets received at each sensor node (in red the landmarks nodes) from data packets coming from the other sensor nodes during an experiment in SensLAB Rennes platform where each sensor node receives until 6000 packets. Given a value in line i and column j we obtain the number of received packets at sensor node i coming from sensor node j.

estimated distances by 8 of the 16 sensor nodes in the network: (214, 176, 178, 225 232, 163 236 and 244). Distances are estimated using the RSSI data received at each sensor node according to the expression of the unbiased estimator of distances (cf. equation (4.11)).

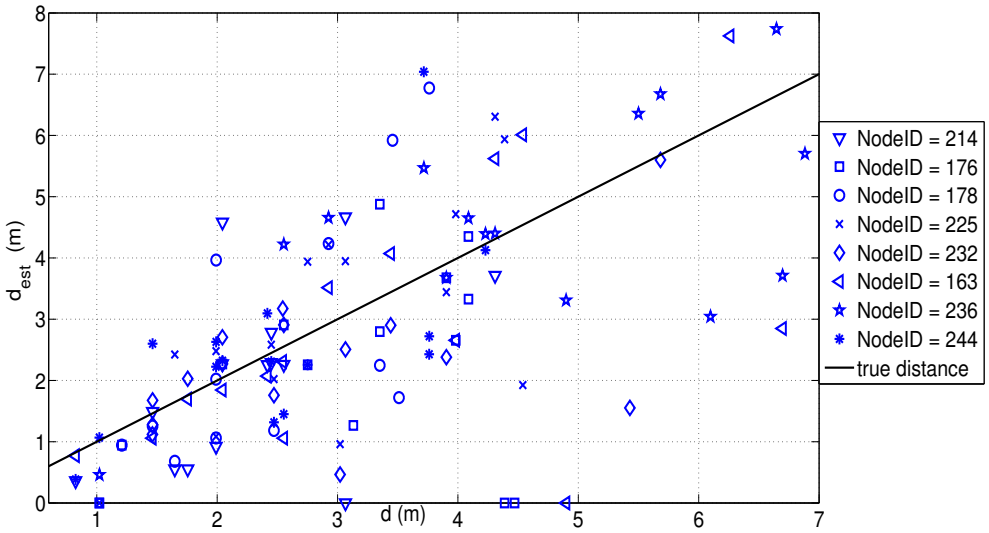


Figure 4.8: Estimated distances $\hat{d}$ (blue markers) using the data received at 8 sensor nodes. The black straight line corresponds to the true distance d.

4.2.2 Standard methods

4.2.2.1 Random localization

Random Localization (RL) process uses statistical Monte Carlo approach to draw random coordinates according to a uniform distribution over the whole room corresponding to each testbed. This random localization process is used to evaluate the contribution of RSSI under the assumption of the Gaussian distribution. A full description of Monte Carlo algorithm is available in [174].

Monte Carlo method is a algorithm considered from a statistical point of view. It consists of generating a large number of independent realizations (set of random samples) under a given distribution. For instance, it can be applied to represent the best distribution of RSSI according to a log-normal shadowing model.

4.2.2.2 Localization mechanism based on geometrical techniques: Trilateration [159, 5, 6, 88]

In this section we recall the main methods used to solve the localization problem from a geometrical point of view. Some of them appeared to be implemented in the GPS localization (*i.e.* trilateration). Such methods are characterized by the need of a minimum number of landmarks to solve the localization problem. Localization mechanism based on a geometrical approach uses a linear equation system. They are based on the principle of circles intersection [88, 1, 20]. Each blind mobile sensor node aiming to compute its position needs to be in the communication range of at least 3 non-collinear landmarks nodes in two-dimensional case. The formal mechanism based on this geometrical approach is the trilateration.

Distance measurements are obtained from the received signals through the measured RSSI and as long as the log-normal shadowing propagation model is assumed. With trilateration, the resulting position of an object depends on the estimation of the distance and the positions of the landmarks. Trilateration uses the geometric properties of access points (or reference nodes)

locations to estimate the location of a blind mobile node. Moreover, due to internal factors disturbances, this technique requires a precise and predictable relationship between distance and the measured RSSI. We defined d_j as a distance between the landmark j and a given blind node willing to localize itself. With trilateration, this distance leads to the equation of a circle whose center is the landmark j. In that case the blind node can estimate the distance d_j through RSSI values measured on data packets sent by the landmark j.

Instead, in the context of min-max which is another localization mechanism based on geometrical technique and presented in appendix F, the estimation of this distance d_j conducts to the equation of a quadrilateral.

Two estimated distances d_j, from two different landmarks will therefore lead to two equations of circles for trilateration method. And so on, as shown in Figure 4.9, three estimated distances from three different landmarks lead to the intersection of three circles. In the ideal case, the intersection of three circles for trilateration leads exactly to the unknown position of the blind mobile node. In the case of an accurate trilateration mechanism, two circles will intersect in two points and the blind mobile node 2D coordinates are determined using a third circle according to a third nearest landmark.

Such geometric-based localization method requires a small number of computations and communications and are effective when landmarks are well placed. In general, emplacement of landmarks are the corners of the surrounding area in which the sensor nodes are located, according to the results in [86, 175].

There are two variants for the trilateration mechanisms: the range measurement techniques (also known as lateration) and the angulation-based techniques. The Range Measurement Techniques or lateration is based on distance measurements. It considers the position of an object by measuring its distance from multiple landmarks/reference points. In two dimensional case ($2D$), you need at least three landmarks/reference points and a target named blind mobile node while 4 landmarks/reference points and at least one blind mobile node are required in three dimensional case (3D). However, as described in many papers, if the antenna array of sensor nodes is above the blind mobile node, two reference points are enough in $2D$ environment and three reference points in a $3D$ one. Distance computation is possible with metrics like the Received Signal Strength Indicator (RSSI), the Time of Arrival (ToA) and its variant. Such localization metrics are presented in Section 2.2. A more general overview can also be found in [176]. Notice in the case of trilateration, the radius r_j of any circle j is supposed to be equal to d_j. Using the known positions (x_j, y_j) of each landmark and the distances r_j, the blind mobile node position results in solving the following systems of three equations in $2D$ or $3D$ case:
$3D$ positioning:

$$\left\{ \begin{array}{lcl} (x_{BN} - x_1)^2 + (y_{BN} - y_1)^2 + (z_{BN} - z_1)^2 & = & r_1^2 \\ (x_{BN} - x_2)^2 + (y_{BN} - y_2)^2 + (z_{BN} - z_2)^2 & = & r_2^2 \\ (x_{BN} - x_3)^2 + (y_{BN} - y_3)^2 + (z_{BN} - z_3)^2 & = & r_3^2 \end{array} \right\}.$$

$2D$ positioning:

$$\left\{ \begin{array}{lcl} (x_{BN} - x_1)^2 + (y_{BN} - y_1)^2 & = & r_1^2 \\ (x_{BN} - x_2)^2 + (y_{BN} - y_2)^2 & = & r_2^2 \\ (x_{BN} - x_3)^2 + (y_{BN} - y_3)^2 & = & r_3^2 \end{array} \right\}.$$

To find the absolute position in three dimensional (3D), one can assume that the centers of the

spheres lie on the $z = 0$ plane (*i.e.* $z_i = 0$ for $i = 1, 2, 3$). A reference coordinate system with the origin at $(0, 0)$ is taken at one of the centers of the circles. In our case, we have chosen the first landmark as the origin of the reference coordinate system and we have chosen the second landmark passing on its X-axis. The position of the other landmarks and regular wireless sensor nodes including all blind mobile nodes are related to this reference system. Hence, the relative positions of the three centers of the three circles can be expressed from this reference coordinate system as: $x'_1 = (0, 0)$, $x'_2 = (x_2, 0)$ and $x'_3 = (x_3, y_3)$

The positions on this reference system are expressed according to a translation t and a rotation matrix R from the absolute system as follows:

$$\begin{gathered}(x'_{BN}, y'_{BN}) = [(x_{BN}, y_{BN}) + t]R \\ \text{with :} \quad t = -(x_1, y_1) \quad , \quad R = \begin{bmatrix}\cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta)\end{bmatrix} \quad \text{et} \quad \theta = \arctan\left(\tfrac{y_2 - y_1}{x_2 - x_1}\right)\end{gathered} \tag{4.15}$$

The above linear system of equations are formulated in this reference coordinate system as:

$$\begin{aligned}&x'^2_{BN} + y'^2_{BN} + z'^2_{BN} = r_1^2 \\ &(x'_{BN} - x'_2)^2 + y'^2_{BN} + z'^2_{BN} = r_2^2 \\ &(x'_{BN} - x'_3)^2 + (y'_{BN} - y_3)^2 + z'^2_{BN} = r_3^2\end{aligned}$$

Then the solution $(x'_{BN}, y'_{BN}, z'_{BN})$ is the relative position to the selected reference coordinate system and is given by:

$$\begin{aligned}x'_{BN} &= \frac{r_1^2 - r_2^2 + x_2^2}{2x_2} \\ y'_{BN} &= \frac{r_1^2 - r_3^2 + x_3^2 + y_3^2}{2j} - \frac{i}{j}x' \\ z'_{BN} &= \pm\sqrt{r_1^2 - x'^2_{BN} - y'^2_{BN}}\end{aligned} \tag{4.16}$$

Considering a two-dimensional case, the absolute position of the initial system (x, y) is obtained from the inverse transform of the relative position (x', y'). The absolute position of the blind mobile node (x_{BN}, y_{BN}) is obtained as: $(x_{BN}, y_{BN}) = (x'_{BN}, y'_{BN})R^t - t$. However note that the ambiguity in the sign of the z coordinate should be resolve when the third dimension is considered.

However, in practice, measurements of three distances (d_1, d_2, d_3) are tainted by error, because of the noise due to the phenomena of indoor disturbance. Consequently, the intersection of the circles result in an area instead of a point, or in some cases there could not be intersection at all. Assuming that the distances are estimated with noise from the RSSI measurements according to the log-normal shadowing model. In Figure 4.9 we illustrate and example with estimated distances from the log-normal shadowing model where $\text{PL}(d_0) = 0dBm$.

Another drawback of this method is the need of the rotation angle θ of the coordinate system. A possible option may be to have an angle of arrival (AOA) measurement between the first two landmarks nodes, which is then sent to the other landmarks in the system. However, these

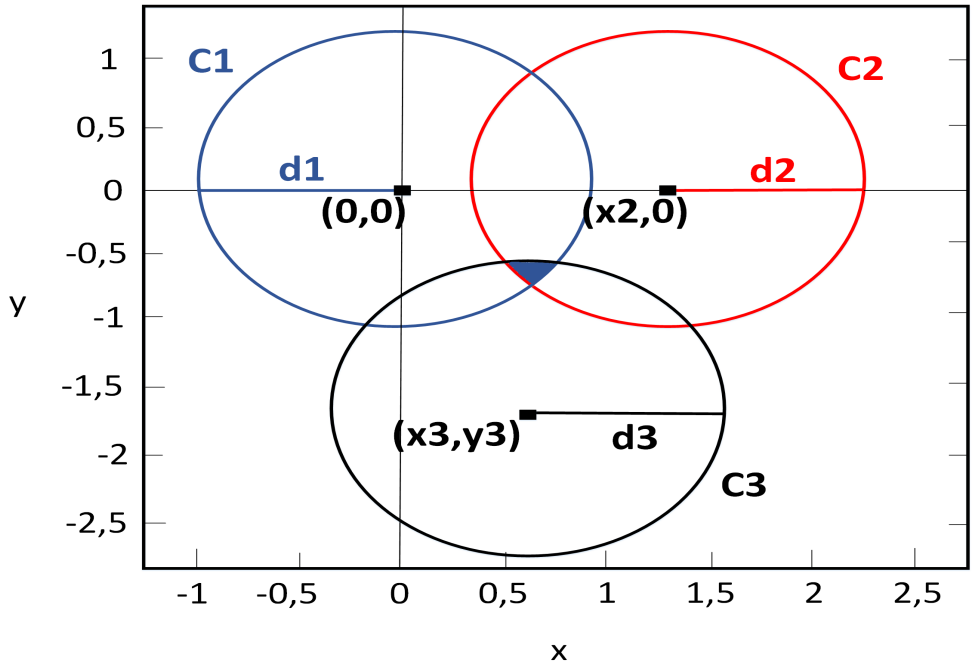


Figure 4.9: Intersection of three circles when considering the center of the first circle as a reference $(0,0)$, its horizontal axis intersecting the center of the second circle: $(x_2,0) = (1.5,0)$. The third circle is centered at the coordinates: $(x_3,y_3) = (0.75,-0.6614)$. Example of a noisy environment in which the intersection point (unknown position of the blind mobile node (x'_{BN},y'_{BN}) lies on the colored area) does not accurately measure the radii (d_1,d_2,d_3) but erroneous distances: $(\hat{d}_1,\hat{d}_2,\hat{d}_3)$.

.

steps are implicit when taking one more landmark as a reference and include its information in the system of equations to be solved. This is considered in the version of trilateration and multilateration we have implemented.

In most cases, in a real environment, a precise measurement of the three distances (r_1,r_2,r_3) are not obtained but a noisy version. Assuming these distances $(\hat{r}_1,\hat{r}_2,\hat{r}_3)$ follow the noisy log-normal shadowing model and estimated from the RSSI measurements with $\mathrm{PL}(d_0) = 0\mathrm{dBm}$; we define:

$$\hat{r}_i = r_i 10^{\xi_i} \quad \text{with :} \quad iid\ \xi_i \sim \mathcal{N}(0, \frac{\sigma^2}{(10\eta)^2})\,,\ i = 1,2,3 \tag{4.17}$$

This expression defined in equation 4.17 holds for any constant α, $E[\hat{r}_i^{\alpha}] = r_i^{\alpha} K^{\alpha^2}$ with $K = E[10^{\xi_i}] = 10^{\ln 10 \sigma_i^2/2}$. The estimated position $(\hat{x}'_{BN},\hat{y}'_{BN})$ of the blind mobile node being in the colored field as presented in Figure 4.9 can be obtained from the expressions (4.16) with $(\hat{r}_1,\hat{r}_2,\hat{r}_3)$.

An estimation of the relative position $(\hat{x}'_{BN},\hat{y}'_{BN})$ is obtained which is generally biased and whose resulting mathematical expectation is:

$$
\begin{aligned}
E[\hat{x}'_{BN}] =& x'_{BN} + (K^4 - 1)\frac{r_1^2 - r_2^2}{2x_2} \\
E[\hat{y}'_{BN}] =& y'_{BN} + (K^4 - 1)(\frac{r_1^2 - r_3^2}{2y_3} - \frac{x_3}{y_3}\frac{r_1^2 - r_2^2}{2x_2})
\end{aligned}
\tag{4.18}
$$

It is observed from the expressions in equation (4.16) that the coordinates are no longer independent each other and since the coordinate $\hat{y}'$ depends on the component $\hat{x}'$, there is a non null correlation of the crossed term $E[(\hat{x}' - E[\hat{x}'])(\hat{y}' - E[\hat{y}'])]$ which appears in the covariance matrix of the estimate.

Note that a position has two coordinates (or components or dimensions) and is represented by two variables that we have denoted x and y. The positions (x, y) is represented in an absolute coordinate system where the origin is $(0, 0)$. As we get estimated values, the estimated position is denoted by its estimated coordinates $\hat{x}$ and $\hat{y}$. Taking one of the landmarks as the origin, we compute a relative estimated position whose estimated coordinates are denoted by $\hat{x}'$ and $\hat{y}'$.

These estimated coordinates, which are often affected by noise, are modeled as random variables characterized by their expectations (thus the bias) and their covariances (the variances when the cross term $E[(\hat{x}' - E[\hat{x}'])(\hat{y}' - E[\hat{y}'])]$ is zero, *i.e* the variables are independent).

Therefore, if we want to estimate the position (two coordinates (x, y)) with the trilateration method represented in a separate orthogonal Cartesian axis (for two independent coordinates), we observe that the values are not represented independently because the dimension of the estimated value $\hat{y}'$ depends on the value of the estimated $\hat{x}'$ dimension or vice versa (if we get first $\hat{y}'$ as a function of $\hat{x}'$). The empirical results have demonstrated that the cross term $E[(\hat{x}' - E[\hat{x}'])(\hat{y}' - E[\hat{y}'])]$ is not null and thus the axes are not orthogonal. For example, in the x direction, if the channel is affected by a noise, it may affect the estimated value for the coordinate y. In the case if one is interested by the y coordinate for some reason, for example when tracking a blind node, the estimated 2D coordinate will be biased. However, trilateration is a simple with low computational cost localization method.

In addition, when considering the expression in equation (4.18) there is a null bias in two cases: for $K \to 1$ (low noise environment $(\xi_i \sim 0)_{\forall i}$) or when the landmarks are equidistant from the blind mobile node with unknown position (that is $r_1 = r_2 = r_3 = r$). In the second case (equidistant one) the equations 4.16 are simplified giving the unknown position coordinates to estimate as $(x', y') = (\frac{x_2}{2}, \frac{x_3^2 + y_3^2 - x_3 x_2}{2y_3})$. Then, the components of the estimated position can be written as a noisy version of the true values such:

$$
\begin{aligned}
\hat{x}' =& x' + \frac{r^2}{2x_2}(10^{2\xi_1} - 10^{2\xi_2}) \\
\hat{y}' =& y' + \frac{r^2}{2y_2}(10^{2\xi_1} - 10^{2\xi_3} - \frac{x_3}{x_2}(10^{2\xi_1} - 10^{2\xi_2}))
\end{aligned}
\tag{4.19}
$$

The covariance matrix computed as $E[(\hat{x}'_{BN} - x'_{BN}, \hat{y}'_{BN} - y'_{BN})^t(\hat{x}'_{BN} - x'_{BN}, \hat{y}'_{BN} - y'_{BN})]$ results

in:

$$\frac{K^8 r^4}{2}(K^8-1)\begin{bmatrix} \frac{1}{x_2^2} & \frac{1}{2y_3x_2}(1-\frac{2x_3}{x_2} \\ \frac{1}{2y_3x_2}(1-\frac{2x_3}{x_2}) & \frac{1}{y_3^2}(1-\frac{x_3}{x_2})+\frac{x_3^2}{x_2^2}) \end{bmatrix} \tag{4.20}$$

These expressions are evaluated using the example in Figure 4.9 with values $r = 1$, $x_2 = 1.5$, $x_3 = 0.58$ and $y_3 = -1.65$ and for different values of σ^2. We illustrated in Figure 4.10 that when considering the trilateration unbiased estimator of the distance, the mean squared error of the estimated positions are nearly the values of the variances defined by the theoretical expressions in (4.20).

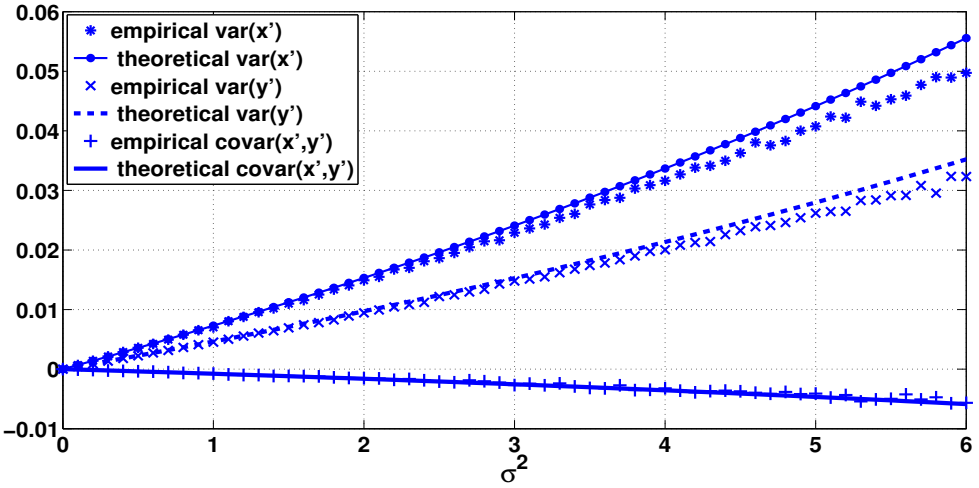


Figure 4.10: Theoretical values of the variances of the two coordinates $\hat{x}'$ and $\hat{y}'$ and its covariance $\hat{x}'-\hat{y}'$ obtained with the expression (4.20); and with their empirical values from 1000 independent Monte Carlo realizations (RSSI data generated with Matlab) for the estimated distances $(\hat{r}_1, \hat{r}_2, \hat{r}_3)$ and when considering different values of σ^2.

We also demonstrated in appendix F that the trilateration localization mechanism is more accurate but less robust against the presence of noise than Min-Max localization algorithm. Note additionally, when there is no intersection between the involved circles according to RSSI measurements from three landmarks when using trilateration, the blind mobile node will not be able to estimate its $2D$ coordinates.

RSSI-based indoor localization using the trilateration approach (named 3later) Below is the trilateration algorithm we implement from the equation of the log-normal shadowing 4.1:

$$d_i = d_0.10^{\frac{\mathrm{PL}(d_0)-\mathrm{PL}(d_i)}{10.\eta}} \tag{4.21}$$

Let us denote by d_i the distance that separates a blind node to any landmark node i with coordinates (x_i, y_i); d_j the distance that separates the same blind node to another landmark j with coordinates (x_j, y_j). We define $R_{i,j}$, the distance ratio between two landmarks nodes i and

j. If d_k is the distance between the blind node and a third landmark k with coordinates (x_k, y_k), we can deduce:

$$d_i = R_{i,j} d_j. \qquad d_i = R_{i,k} d_k \tag{4.22}$$

Note that the equation $d_i = R_{j,k} d_k$ can be easily obtained from the above two relations defined in the equation 4.22.

Replacing d_i, d_j and d_k by their expression of euclidean distance and by squaring: $d_i^2 = R_{i,j}^2 d_j^2$ and $d_i^2 = R_{i,k}^2 d_k^2$, we obtain the following two forms of equations of two conical:

$(x-x_i)^2+(y-y_i)^2) = R_{i,j}^2(x-x_j)^2+R_{i,j}^2(y-y_j)^2$ $(x-x_i)^2+(y-y_i)^2) = R_{i,k}^2(x-x_k)^2+R_{i,k}^2(y-y_k)^2$

By developing the expressions of the conical:

$x^2 - 2xx_i + x_i^2 + y^2 - 2yy_i^2 + y_i^2 - R_{i,j}^2 x^2 + 2R_{i,j}^2 xx_j - R_{i,j}^2 x_j^2 - R_{i,j}^2 y^2 + 2R_{i,j}^2 yy_j - R_{i,j}^2 y_j^2 = 0$

$x^2 - 2xx_i + x_i^2 + y^2 - 2yy_i^2 + y_i^2 - R_{i,k}^2 x^2 + 2R_{i,k}^2 xx_k - R_{i,k}^2 x_k^2 - R_{i,k}^2 y^2 + 2R_{i,k}^2 yy_k - R_{i,k}^2 y_k^2 = 0$

then

$(1 - R_{i,j}^2)x^2 + (1 - R_{i,j}^2)y^2 + 2(R_{i,j}^2 x_j - x_i)x + 2(R_{i,j}^2 y_j - y_i)y + x_i^2 + y_i^2 - R_{i,j}^2 x_j^2 - R_{i,j}^2 y_j^2 = 0$

$(1 - R_{i,k}^2)x^2 + (1 - R_{i,k}^2)y^2 + 2(R_{i,k}^2 x_k - x_i)x + 2(R_{i,k}^2 y_k - y_i)y + x_i^2 + y_i^2 - R_{i,k}^2 x_k^2 - R_{i,k}^2 y_k^2 = 0$

Let us mention:

$A_{i,j} = (1 - R_{i,j}^2)$, $B_{i,j} = 0$, $C_{i,j} = (1 - R_{i,j}^2)$, $D_{i,j} = 2(R_{i,j}^2 x_j - x_i)$, $E_{i,j} = 2(R_{i,j}^2 y_j - y_i)$ and $F_{i,j} = x_i^2 + y_i^2 - R_{i,j}^2 x_j^2 - R_{i,j}^2 y_j^2$

Finally we have:

$$A_{i,j}x^2+B_{i,j}xy+C_{i,j}y^2+D_{i,j}x+E_{i,j}y+F_{i,j} = 0 \qquad A_{i,k}x^2+B_{i,k}xy+C_{i,k}y^2+D_{i,k}x+E_{i,k}y+F_{i,k} = 0 \tag{4.23}$$

We recognize two conical equations like ellipse, hyperbola or parabola. We must then determine the intersection of two conical which may give 0, 1, 2, 3 or 4 intersection points [177, 178].

In the case when two intersection point are obtained, a third distance estimation between the same blind mobile node and a fourth landmark is needed to find the unique intersection point, that is the position of the blind mobile node. In fact the unique intersection point, if exist, will be one of the two points (from the conical intersections) that is closer to the third conical defined by the distance estimate d_l between the blind mobile node and a fourth landmark l. Figure 4.11 presents the pseudo code of the trilateration algorithm assuming $n = 4$ landmarks:

With any number of landmarks in the localization system, we will always choose 4 distances measurements between the mobile node and 4 different landmarks among n landmarks in total to find a unique intersection point at each step of the iteration of the trilateration algorithm until the algorithm converges. This recursive version of trilateration that we name *Lodachai* first find the intersection of the two conical C_1, C_2 as a result of three distances measurements (two distances ratio) from three different landmarks. This first intersection of the two conical results either in two intersection points $p_{1,2}$ and $p_{1,2}\prime$ or in one intersection point $p_{1,2}$ if $d(C_1, C_2) \leq r_1 + r_2$. $d(C_1, C_2)$ is the distance to both centers of both conical if $R_{1,2} \neq 1$ and $R_{1,3} \neq 1$ and $r_1 + r_2$ are the radii of the conical. In the case for one intersection, this value is retained and stored in the list of intersection points. Whereas, if the two conical result in two intersections points, the

$n \geq 4$
For i = 1 to n-3 do
For j = i+1 to n-2 do
For k = i+2 to n-1 do

compute $R_{i,j} = 10^{\frac{(PL(d_i)-PL(d_j))}{10\eta}}$

compute $R_{i,k} = 10^{\frac{(PL(d_i)-PL(d_k))}{10\eta}}$

compute $A_{i,j} = 1 - {R_{i,j}}^2, B_{i,j} = 0, C_{i,j} = 1 - {R_{i,j}}^2, D_{i,j} = 2({R_{i,j}}^2 x_j - x_i),$
$E_{i,j} = 2({R_{i,j}}^2 y_j - y_i), F_{i,j} = x_i^2 + y_i^2 - {R_{i,j}}^2 x_j^2 - {R_{i,j}}^2 y_j^2$

compute $A_{i,k} = 1 - {R_{i,k}}^2, B_{i,k} = 0, C_{i,k} = 1 - {R_{i,k}}^2,$
$D_{i,k} = 2({R_{i,k}}^2 x_k - x_i), E_{i,k} = 2({R_{i,k}}^2 y_k - y_i),$
$F_{i,k} = x_i^2 + y_i^2 - {R_{i,k}}^2 x_k^2 - {R_{i,k}}^2 y_k^2$

compute the intersection of the equations of two conical

$A_{i,j}x^2 = B_{i,j}xy + C_{i,j}y^2 + D_{i,j}x + E_{i,j}y + F_{i,j} = 0$

$A_{i,k}x^2 = B_{i,k}xy + C_{i,k}y^2 + D_{i,k}x + E_{i,k}y + F_{i,k} = 0$

If one intersection points
Select it
Else if two intersections points
For l = i + 3 to n do
Find the intersection point that is closest to l
End
Else
There is no solution
End
Retain the nearest intersection point to the conical l
End
End
End
Compute the centroid of the selected intersection points

Figure 4.11: The pseudo code of the trilateration algorithm

retained point is the one that is the most closest to a third conical defined by the distance ratio: $d_i = R_{i,l} d_l$.

This method of finding a solution (intersection) for a given location of a blind node in a localization system in indoor with n landmarks is done iteratively and verifies all possible intersection points. This heuristic associates at each iteration 4 among n landmarks chosen according $\binom{n}{4}$. For instance, if $n = 5$ landmarks, the solution leading to the blind node position is the centroid of the intersection points determined with quadruplets (LM_1, LM_2, LM_3, LM_4), (LM_1, LM_2, LM_3, LM_5), (LM_1, LM_2, LM_4, LM_5), (LM_1, LM_3, LM_4, LM_5), (LM_2, LM_3, LM_4, LM_5). Finally, the centroid of all the intersection points is taken as the estimated position of the blind node by summing all individual previously retained intersections and dividing by their total count.

It is easy to demonstrate that the points PT that verify the equation $d(PT, PT_i) = R_{i,j} d(PT, PT_j)$ is an ellipse and more precisely a circle of center C $(\frac{(R_{i,j}^2 X_j - X_i)}{(R_{i,j}^2 - 1)}, \frac{(R_{i,j}^2 Y_j - Y_i)}{(R_{i,j}^2 - 1)})$ and radius $(\frac{R_{i,j} * d(P_i, P_j)}{\sqrt{(R_{i,j}^2 - 1)^2}})$: (cf. appendix D).

Clearly, trilateration involves the distances measurements from at least three landmarks. Meanwhile the multilateration [88] is the version of the trilateration involving RSSI measurements, then distance estimation from at least four different landmarks nodes. Instead, the multiangulation uses the same principle as multilateration but with angle measurements from at least four landmarks.

Multilateration Multilateration is another method based on the geometric analysis and is an extension of the trilateration case using the cosine theorem. As presented in Figure 4.12, multilateration measures the difference between the square of the distances and conduct to a number of locations that satisfy the distance measurements. This is different to trilateration or triangulation that respectively use the distances and angles measurements. If we have M landmarks (sensor nodes with known positions) surrounding a sensor node, the cosine theorem is applied to any of the $M - 1$ triangles formed from three sensor vertices. Each triangle corresponds to : a landmark chosen as a reference, one of the each remaining $M - 1$ landmark and the sensor node itself. We can apply this method to a set of N sensor nodes having a set of M available landmarks for each sensor node.

Then, each of the N regular sensor nodes obtains its position independently using the data from M landmarks. Therefore, with one landmark chosen as a reference, each regular node has $M-1$ equations to solve to estimate its position.

Assume a blind mobile node receiving the positions of M landmarks nodes $\{(x_i, y_i)\}_{i=1,...,M}$ and getting at the same time the RSSI measurements to estimate the distances separation between itself and each landmark. The purpose of this method is to construct a linear system of $M - 1$ equations with the two unknowns coordinates (x_{BN}, y_{BN}). The solution for the equations is obtained by solving a linear optimization problem: minimizing the mean square error associated to distance measurements.

Consider a triangle of three vertices $\{S_1, S_2, S_3\}$=$\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$, for instance like the one presented in Figure 4.13 whose distances $\{d_{12}, d_{13}, d_{23}\}$ correspond to the three sides and γ being the angle of one vertex let say (x_1, y_1); the cosine law is applied as follows: the square of the length of a side of any triangle is equal to the sum of the squares of the two other sides minus twice the product of these sides by the cosine of the angle between these two sides.

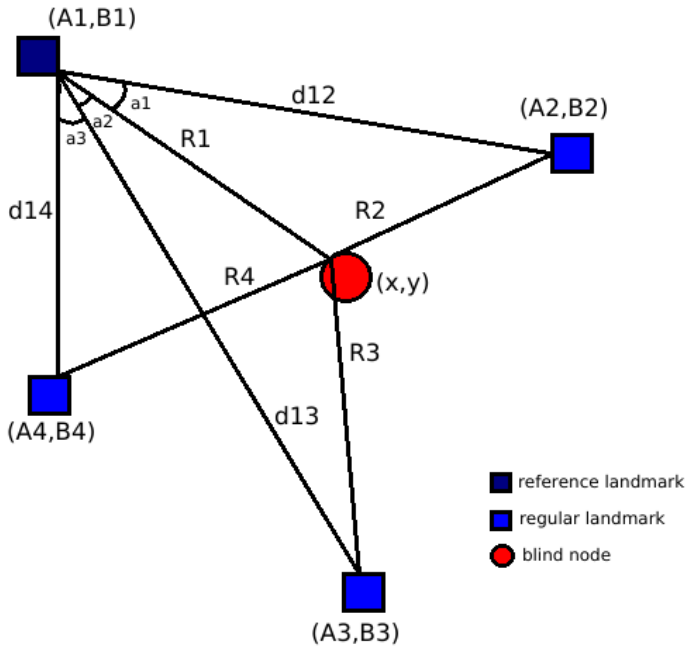


Figure 4.12: Overview of the multilateration approach

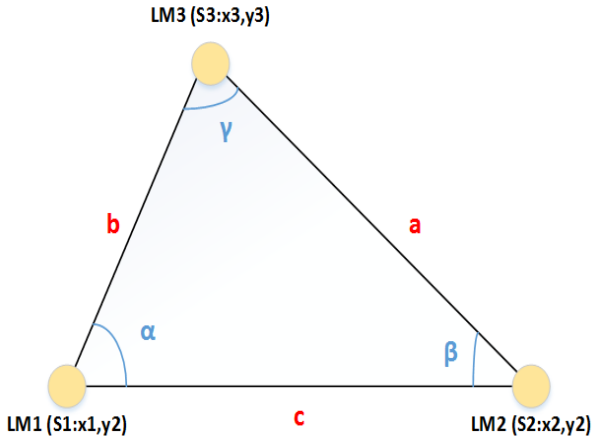


Figure 4.13: The cosinus law says: $c^2 = a^2 + b^2 - 2ab\cos\gamma$. The formula is applied to our problem statement with the sides (a, b, c) corresponding to sensor node distances (d_{23}, d_{12}, d_{13}).

$$d_{23}^2 = d_{12}^2 + d_{13}^2 - 2d_{12}d_{13}\cos\gamma = d_{12}^2 + d_{13}^2 - 2d_{12}d_{13}\frac{< \overrightarrow{S_1S_2}, \overrightarrow{S_1S_3} >}{d_{12}d_{13}}$$
$$d_{23}^2 = d_{12}^2 + d_{13}^2 - 2\left((x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_1)\right)$$

If this cosine law is applied to the $M-1$ triangles formed by $\{(x_1, y_1), (x, y), (x_i, y_i)\}$ with $i = 2, \ldots, M$. The $M-1$ equations are obtained from the previous expression with (x_1, y_1) being the position of the first landmark node used as a reference for all triangles, (x_{BN}, y_{BN}) being the unknown position of the blind mobile node to estimate and the remaining positions correspond to those of the $M-1$ remaining landmarks. The system is written as:

$$
\begin{aligned}
&d_1^2 + d_{12}^2 - 2\left((x_{BN} - x_1)(x_2 - x_1) + (y_{BN} - y_1)(y_2 - y_1)\right) = d_2^2 \\
&d_1^2 + d_{13}^2 - 2\left((x_{BN} - x_1)(x_3 - x_1) + (y_{BN} - y_1)(y_3 - y_1)\right) = d_3^2 \\
&d_1^2 + d_{14}^2 - 2\left((x_{BN} - x_1)(x_4 - x_1) + (y_{BN} - y_1)(y_4 - y_1)\right) = d_4^2 \\
&\vdots \\
&d_1^2 + d_{1M}^2 - 2\left((x_{BN} - x_1)(x_M - x_1) + (y_{BN} - y_1)(y_M - y_1)\right) = d_M^2
\end{aligned}
$$

We rewrite the above expression under a matrix form in order to see it as a linear system $\{d_i\}_{i=1,\ldots,M}$ with distances between the blind mobile node and the M landmarks nodes:

$$
\boldsymbol{b} = A\boldsymbol{\theta} \quad \text{with :} \quad \boldsymbol{b} = \boldsymbol{d} - \boldsymbol{\delta} = \frac{1}{2}\begin{bmatrix} d_{12}^2 - (d_2^2 - d_1^2) \\ \vdots \\ d_{1M}^2 - (d_M^2 - d_1^2) \end{bmatrix}
$$

$$
A = \begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) \\ \vdots & \vdots \\ (x_M - x_1) & (y_M - y_1) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\theta} = (x_{BN}, -x_1\ ,\ y_{BN}, -y_1)^t = \boldsymbol{z} - (x_1, y_1)^t \qquad (4.24)
$$

Note that the variable t within the equation 4.24 corresponds to the mathematical transpose formula. This geometric method gives a system of equations whose vector form is $\boldsymbol{b} = A\boldsymbol{\theta}$, with $\boldsymbol{\theta}$ dependent on the unknown position $\boldsymbol{z}$ to compute. Note that this system of equation 4.24 corresponds to an ideal situation without noise. Such system also considers that all involved values can be known that is the distances and thus positions. However, the vector $\boldsymbol{b}$ is not known in our situation, we obtain a noisy (estimated) version denoted by $\hat{\boldsymbol{b}}$. Next, we consider on one hand having the vector with the square of the distance between landmarks $(d_{12}^2, d_{13}^2, \ldots, d_{1M}^2)$, that is in general $\{d_{1i}^2\}_{i=2,\ldots,M}$.

This vector of distances is well known so we do not add the notation hat ($\hat{}$) above the corresponding variable. Recall that from a statistical point of view, the notation hat ($\hat{}$) above a variable is always for the notion of estimation, *i.e.*, when trying to calculate the best possible way a value of a variable that is unknown, noisy. This variable is always a random variable whose average (which implies to be the true value, so no bias) and variance can be computed; or in the case of a vector, we may calculate the covariance matrix that is the best we can estimate. This covariance matrix is expected to be a diagonal matrix whose variances are the smallest. On the other hand there is a vector δ that contains the squares of the distances between the blind node to localize and the M landmarks $(d_1^2, \ldots, d_M^2)$ in the mentioned system of equations (cf. equation 4.24) given by the geometric method. However, this vector is not known due to the presence of noise on measured RSSI values; we denote $\hat{\boldsymbol{\delta}}$ as the unbiased estimate of the vector $\boldsymbol{\delta}$ (using the square of the distance estimator $\hat{d}_{square}$ already presented in equation 4.13). Thus, a noisy version (unbiased estimated) denoted $\hat{\boldsymbol{b}} = \boldsymbol{d} - \hat{\boldsymbol{\delta}}$ is obtained and involved in the system of equations defined in 4.24. Then, we try to minimize the mean square error between the real value of the vector $\boldsymbol{b}$ equal to $A\boldsymbol{\theta} = A(\boldsymbol{z} - (x_1, y_1))$ and the obtained noisy vector $\hat{\boldsymbol{b}} = (\boldsymbol{d} - \hat{\boldsymbol{\delta}})$ which is also the one that is unknown and can be measured through the RSSI values.

To summarize , the norm $||\hat{\boldsymbol{b}} - \boldsymbol{b}||^2 = ||\hat{\boldsymbol{b}} - A\boldsymbol{\theta}||^2 = ||\hat{\boldsymbol{b}} - A\boldsymbol{\theta}||^2 = ||\hat{\boldsymbol{b}} - A(\boldsymbol{z} - (x_1, y_1))||^2$ is minimized. $\hat{\boldsymbol{b}}$ is measured from $\boldsymbol{d}$ and $\hat{\boldsymbol{\delta}}$; the matrix A is computed, the position (x_1, y_1) of a reference landmark is known, and therefore, we aim to compute the unknown position $\boldsymbol{z}$ of the blind node. Then the best estimate for $\boldsymbol{z}$ is found and it is denoted by $\hat{\boldsymbol{z}}$. The system of linear equations is solved by a *least squares* method since the distribution of the resulting equations is not known.

Using the M known landmarks positions $(x_i, y_i)_{i=1,...,M}$, the vector $\boldsymbol{d}$ contains the set of the squares distances $\{d_{1i}^2\}_{i=2,...,M}$ between the reference landmark 1 and the remaining $M-1$ landmarks. The vector $\boldsymbol{\delta}$ contains the $M-1$ differences between the square of the distances at the blind node $d_i^2 - d_{1_{i=2,...,M}}^2$ from the reference landmark 1 and the $M-1$ remaining landmarks.

As the vector $\boldsymbol{\delta}$ is not perfectly known by the blind node, instead, it obtains estimation of the set of $\{\hat{d}_i^2\}_{i=1,...,M}$ received from the M landmarks. In such case, we denote as $\hat{\boldsymbol{\delta}}$ it resulting noisy version and we seek to find the estimated position of the blind node $\hat{\boldsymbol{z}}$ that minimizes the mean square error between the real and the estimated vector computed as the linear system described above. The problem is summarized as follows :

$$\hat{\boldsymbol{z}} = \min_{\boldsymbol{z}} ||\hat{\boldsymbol{b}} - A\boldsymbol{\theta}||^2 = \min_{\boldsymbol{z}} \left(\hat{\boldsymbol{b}}^t \hat{\boldsymbol{b}} + \boldsymbol{\theta}^t A^t A \boldsymbol{\theta} - 2\hat{\boldsymbol{b}}^t A\boldsymbol{\theta} \right) \tag{4.25}$$

The solution of the estimated position of the blind mobile node is obtained:

$$\hat{\boldsymbol{z}} = C(\boldsymbol{d} - \hat{\boldsymbol{\delta}}) + (x_1, y_1)^t \tag{4.26}$$

while defining the matrix $C = (A^t A)^{-1} A^t$.

As multilateration is a generalization of the trilateration method, it provides absolute positions at one step without needing a last step to calculate the rotation and translation of the principal axes. In fact, these steps are implicit when taking one more landmark as the reference and including its position in the $M-1$ linear system of equations. While for trilateration, M is equal to 3 for two dimensional case, multilateration needs at least 4 landmarks. With multilateration, several blind nodes can simultaneously compute their positions.

Since the expression of $\hat{\boldsymbol{z}}$ depends on the estimations of the square of the distances $\{\hat{d}_i^2\}_{i=1,...,M}$ the solution of equation 4.25 results an unbiased estimate of the position as long as the unbiased ML estimator $\hat{d}_{squared}$ defined in equation 4.9 is used. Hence, the multilateration method gives a random estimated position $\hat{\boldsymbol{z}}$ such $E[\hat{\boldsymbol{z}}] = \boldsymbol{z}$ and its covariance matrix $\text{cov}[\hat{\boldsymbol{z}}]$ is computed as :

$$\begin{aligned} E[(\hat{\boldsymbol{z}} - E[\hat{\boldsymbol{z}}])(\hat{\boldsymbol{z}} - E[\hat{\boldsymbol{z}}])^t] &= C\text{cov}[\hat{\boldsymbol{\delta}}]C^t \\ \text{where :} \quad \text{cov}[\hat{\boldsymbol{\delta}}] &= (K^8 - 1) \begin{bmatrix} d_1^4 + d_2^4 & d_1^4 & \cdots & d_1^4 \\ d_1^4 & d_1^4 + d_3^4 & \vdots & \\ \vdots & \ddots & \vdots & \\ d_1^4 & d_1^4 & \cdots & d_1^4 + d_M^4 \end{bmatrix} \end{aligned} \tag{4.27}$$

We observe that estimated position by multilateration has no bias, as the unbiased maximum likelihood estimator $\hat{d}_{square}$ is used for all the square distances separation from the blind node to the set of landmarks. The resulting unbiased position $\hat{z}$ has a covariance matrix defined in equation 4.27 which depends on the noise level (as long as constant K contains σ^2) and the corresponding distances with exponent 4.

Thus, we may note from the above expression three main facts : firstly, the variance of the estimated position by the multilateration method depends proportionally on the noise level σ^2 throughout the factor K^8 and the distances to the M landmarks (far distant landmarks are not desired); secondly, in general the estimated position has correlated coordinates since the

resulting covariance matrix is not a diagonal matrix and the cross terms are not in general equal to zero (having a coordinate not well estimated could negatively affect the remaining estimated coordinate); and finally, the covariance could be null in some cases as long as the factor K tends to one (e.g. low noise level).

To minimize this effect, *i.e.* the ependency between the variance of the resulting estimated position and the distances to the landmarks, one may take into account the noise by introducing in the equation 4.25 a diagonal matrix W of dimensions $(M-1)\times(M-1)$ whose weights are inversely proportional to the computed variance of $\hat{\boldsymbol{\delta}}$. For example, defining W as $\{W(i,i) = 1/(\hat{d}_i^4 + \hat{d}_1^4)\}_{i=2,\dots,N}$, the solution becomes : $\hat{\boldsymbol{z}} = C(\boldsymbol{d} - \hat{\boldsymbol{\delta}}) + (x_1, y_1)^t$ with $C = (A^t W A)^{-1} A^t W$.

The performance of multilateration method is shown more in detail in Section 5.2 using the experimental data acquired from the SensLAB Rennes platform and selecting N blind sensor nodes. These blind sensor nodes form a connected network whose aim is to find its own position using data from 4 landmarks. As it is described later in this chapter, we selected $N = 12$ blind nodes and computed each estimated position (for each blind node) as defined in equation 4.25 from estimated distances according to T RSSI data packets received from the 4 landmarks. We note that such method involves for the whole network a computational cost scaling a factor $N(M-1)$ and a communication cost scaling NMT.

4.2.3 Maximum likelihood optimization-based localization mechanism

The maximum likelihood has been used in many works [5], [179], [164] and is demonstrated to be an accurate and powerful approach to be use in RSSI-based one-hop localization. However other works request that such an approach-based optimization is not a good positioning mechanism [6] as its performance is sensitive to the number of reference nodes.

Maximum Likelihood Estimator (MLE) principle The Maximum Likelihood Estimator is a statistical approach that maximizes the likelihood between the probability density function (pdf) induced by the assumed parametrical model (in our case we assumed data follows a log-Normal distribution) and the pdf function induced by the distribution of the measured data. It is used to find the parameters that maximize the probability of measured data to fit the assumed model. In our context, measured data corresponds to the measured RSSI packets by a set of N sensor nodes in order to find their unknown positions. Hence, the problem states as : find the set of N unknown positions that maximize the likelihood of the measured RSSI data whose distribution is assumed to be Gaussian depending on the parameter set $\boldsymbol{\theta}$. Note that a position (x, y) is implicitly related to the distance parameter d while assuming for example the euclidean distance function $d(x, y)$.

Consider a network of N sensor nodes: $\mathcal{C} = \{1, \dots, N\}$, and the goal is to calculate the set of unknown positions $\{(x_c, y_c)\}_{c\in\mathcal{C}}$ of such nodes. Suppose the signal received by each regular sensor node follows the log-normal shadowing propagation model (cf. equation 4.1). Each regular sensor node $c \in \mathcal{C}$ obtains M RSSI measures from its neighbors $\mathcal{K}_c = \{k \in \mathcal{C} \,|\, k \sim c\}$. Noted as $\boldsymbol{r}_{c,k} := (r(1)_{c,k}, \dots, r(M)_{c,k})$ the vector of RSSI measured at the regular sensor node c from its neighbor k. Assume that the M random variables are independent and identically distributed and follow a Gaussian distribution

$r(m)_{c,k} \sim \mathcal{N}(\, -PL(d_0)_{c,k} - \eta_{c,k} 10\log_{10} d_{c,k}\ ,\ \sigma^2_{c,k})$. If the Gaussian vectors $\{\boldsymbol{r}_{c,k}\}_{k\in\mathcal{K}_c}$ are also independent each other, the joint probability density function of the sensor c can be written as:

$$f_{\theta_c}(\boldsymbol{r}_c) = \prod_{k \in \mathcal{K}_c} f(\boldsymbol{r}_{c,k}) \quad \text{which:} \quad f_{\theta_c}(\boldsymbol{r}_{c,k}) = \frac{1}{(2\pi\sigma^2_{c,k})^{M/2}} e^{-\frac{1}{2\sigma^2_{c,k}} ||\boldsymbol{r}_{c,k} + PL(d_0)_{c,k} + \eta_{c,k} 10 \log 10 d_{c,k}||^2} \tag{4.28}$$

The expression (4.28) is the likelihood function of the parametric model, which depends on all $\theta_c = \{d_{c,k}, PL(d_0)_{c,k}, \eta_{c,k}, \sigma^2_{c,k}\}_{k \in \mathcal{K}_c}$.

Moreover, observations are assumed to be independent among the N regular sensor nodes, the maximum likelihood is expressed aggregating the logarithm of the N pdf functions defined in equation (4.28):

$$\max_{\theta = \{\theta_c\}} \sum_{c=1}^{N} \sum_{k \in \mathcal{K}_c} \mathcal{L}(\theta_c)$$

with :

$$\mathcal{L}(\theta_c) = \ln f_{\theta_c}(\boldsymbol{r}_{c,k}) = -\frac{M}{2} \ln 2\pi - \frac{M}{2} \ln \sigma^2_{c,k} - \frac{1}{2\sigma^2_{c,k}} \sum_{m=1}^{M} \left(r(m)_{c,k} + PL(d_0)_{c,k} + \eta_{c,k} 10 \log 10 d_{c,k} \right)^2 \tag{4.29}$$

The parameter that defines the distance between a regular sensor node c and a regular sensor node k from the equation 4.29 is the euclidean distance, $d_{k,c}$ and is equal to $d_{k,c} = \sqrt{(x_c - x_k)^2 + (y_c - y_k)^2}$. For simplicity, suppose the case $M = 1$, which means one RSSI values is measured at each regular sensor node from data sent by the landmarks for example. The N positions of each blind node is estimated by combining all $P = \frac{N(N-1)}{2}$ measured RSSI values.

$$\boldsymbol{r} := (r(1), \ldots, r(P)) = (r_{1,2}, \ldots, r_{1,n}, r_{2,3}, \ldots, r_{2,n}, \ldots, r_{n-1,n})^T \quad \text{where :} \quad r(p) = r_{i,j} = r(d_{i,j})$$

Following the model of observation (cf. equation 4.28) applied to P values of observation $\boldsymbol{r}$, position estimation by maximum likelihood as described in equation 4.29 corresponds to find the N pairs of values $\theta = ((x_1, y_1), \ldots, (x_N, y_N))$ which satisfy:

$$\hat{\theta} = \max_{\theta} \ \mathcal{L}(\theta)$$

with :

$$\mathcal{L}(\theta) = -\frac{P}{2} \ln 2\pi - \frac{1}{2} \sum_{p=1}^{P} \ln \sigma^2_p + \frac{1}{2\sigma^2_p} \left(r(p) + PL(d_0)_p + \eta_p 10 \log_{10} d_p \right)^2$$

This global optimization problem with $2N$ variables is determined by calculating the gradient $\mathcal{L}(\theta)$ with N partial derivatives in x and y. Note that each pair of variables (x_i, y_i) appears in $N - 1$ terms of the function $\mathcal{L}(\theta)$ and the derivative of the distance function satisfies $\frac{d}{dx_i} d_{i,j} = -\frac{d}{dx_j} d_{i,j}$, then the system of equations is written as:

$$
\begin{aligned}
\tfrac{d}{dx_1}\mathcal{L} &= (r_{1,2} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,2})\tfrac{x_1-x_2}{d_{1,2}^2} + \cdots + (r_{1,n} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,n})\tfrac{x_1-x_n}{d_{1,n}^2} = 0\\
\tfrac{d}{dx_2}\mathcal{L} &= -(r_{1,2} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,2})\tfrac{x_1-x_2}{d_{1,2}^2} + (r_{2,3} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,3})\tfrac{x_2-x_3}{d_{2,3}^2} + \ldots\\
&\quad \cdots + (r_{2,n} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,n})\tfrac{x_2-x_n}{d_{2,n}^2} = 0\\
&\vdots\\
\tfrac{d}{dx_{n-1}}\mathcal{L} &= -(r_{1,n-1} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,n-1})\tfrac{x_1-x_{n-1}}{d_{1,n-1}^2} -\\
&\quad (r_{2,n-1} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,n-1})\tfrac{x_2-x_{n-1}}{d_{2,n-1}^2} - \ldots\\
&\quad \cdots + (r_{n-1,n} + PL(d_0)_{n-1} + \eta_{n-1} 10\log_{10} d_{n-1,n})\tfrac{x_{n-1}-x_n}{d_{n-1,n}^2} = 0\\
\tfrac{d}{dx_n}\mathcal{L} &= -\textstyle\sum_{i=1}^{n-1} \tfrac{d}{dx_i}\mathcal{L} = 0
\end{aligned}
$$

By symmetry, we have the same equations for the partial derivatives in y_i:

$$
\begin{aligned}
\tfrac{d}{dy_1}\mathcal{L} &= (r_{1,2} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,2})\tfrac{y_1-y_2}{d_{1,2}^2} + \cdots + (r_{1,n} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,n})\tfrac{y_1-y_n}{d_{1,n}^2} = 0\\
\tfrac{d}{dy_2}\mathcal{L} &= -(r_{1,2} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,2})\tfrac{y_1-y_2}{d_{1,2}^2} + (r_{2,3} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,3})\tfrac{y_2-y_3}{d_{2,3}^2} + \ldots\\
&\quad \cdots + (r_{2,n} + PL(d_0)_2 + \eta_2 10\log 10 d_{2,n})\tfrac{y_2-y_n}{d_{2,n}^2} = 0\\
&\vdots\\
\tfrac{d}{dy_{n-1}}\mathcal{L} &= -(r_{1,n-1} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,n-1})\tfrac{y_1-y_{n-1}}{d_{1,n-1}^2} -\\
&\quad (r_{2,n-1} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,n-1})\tfrac{y_2-y_{n-1}}{d_{2,n-1}^2} - \ldots\\
&\quad \cdots + (r_{n-1,n} + PL(d_0)_{n-1} + \eta_{n-1} 10\log_{10} d_{n-1,n})\tfrac{y_{n-1}-y_n}{d_{n-1,n}^2} = 0\\
\tfrac{d}{dy_n}\mathcal{L} &= -\textstyle\sum_{i=1}^{n-1} \tfrac{d}{dy_i}\mathcal{L} = 0
\end{aligned}
$$

A system of $2N-2$ equations are obtained with $2N$ variables. If all terms $d_{i,j}$ are always positives and assuming that all positions are different (*i.e.* $x_1 \neq x_2 \cdots \neq x_n$ and $y_1 \neq y_2 \cdots \neq y_n$), P equations are obtained and those are the same equations in both coordinates x and y and must verify the following system of equations:

$$
\begin{aligned}
&(r_{1,2} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,2}) = 0 \quad \Longrightarrow \quad \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2} = 10^{\frac{-r_{1,2}-PL(d_0)_1}{10\eta_1}}\\
&\vdots\\
&(r_{1,n} + PL(d_0)_1 + \eta_1 10\log_{10} d_{1,n}) = 0 \quad \Longrightarrow \quad \sqrt{(x_1-x_n)^2 + (y_1-y_n)^2} = 10^{\frac{-r_{1,n}-PL(d_0)_1}{10\eta_1}}\\
&(r_{2,3} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,3}) = 0 \quad \Longrightarrow \quad \sqrt{(x_2-x_3)^2 + (y_2-y_3)^2} = 10^{\frac{-r_{2,3}-PL(d_0)_2}{10\eta_2}}\\
&\vdots\\
&(r_{2,n} + PL(d_0)_2 + \eta_2 10\log_{10} d_{2,n}) = 0 \quad \Longrightarrow \quad \sqrt{(x_2-x_n)^2 + (y_2-y_n)^2} = 10^{\frac{-r_{2,n}-PL(d_0)_2}{10\eta_2}}\\
&\vdots\\
&(r_{n-1,n} + PL(d_0)_{n-1} + \eta_{n-1} 10\log_{10} d_{n-1,n}) = 0\\
&\Longrightarrow \quad \sqrt{(x_{n-1}-x_n)^2 + (y_{n-1}-y_n)^2} = 10^{\frac{-r_{n-1,n}-PL(d_0)_{n-1}}{10\eta_{n-1}}}
\end{aligned}
$$

These above equations are a system with $P = \frac{N(N-1)}{2}$ equations and $2N$ unknown variables (*i.e.* all N x coordinates and N y coordinates) with the number of measured RSSI values $P \geq 2N$ if the number of involved sensor nodes verifies $N \geq 5$.

As a consequence of such a difficult system, it is not convenient to solve this problem as trying to find the whole N unknown positions at once and without including the landmark nodes. Instead, one can separately resolve the ML estimator at once for each of the blind node position with the only use of RSSI data coming from multiple landmarks nodes named reference nodes.

Then, the location estimation problem with the maximum likelihood estimator can be analyzed by computing the position (x, y) of a sensor node using a set of known positions $(\{(a_m, b_m)\}_{\forall m})$ and K measures $(\{r(m,k)\}_{\forall m,k})$ of M landmarks nodes. In this case, the involved number of communications scales KM for any blind sensor node. Furthermore, the number of operations needed by the blind node when estimating its position with MLE, depends on which algorithm is used to solve this stochastic optimization problem of the maximum likelihood. Such algorithm are explained in the paragraph below. The latter problem is reduced to find the position of the blind node (x_{BN}, y_{BN}) such:

$$(\hat{x}, \hat{y}) = \max_{(x,y)} \ \mathcal{L}(x, y)$$

with :
$\mathcal{L}(x, y) = -\frac{KM}{2} \ln 2\pi - \sum_{m=1}^{M} \sum_{k=1}^{K} \frac{1}{2} \ln \sigma_m^2 + \frac{1}{2\sigma_m^2} \left(r(m,k) + PL(d_0)_m + \eta_m 10 \log_{10} d_m\right)^2$
with $d_m = \sqrt{(x - a_m)^2 + (y - b_m)^2}$

To implement a localization system-based maximum likelihood estimator, multiple algorithms could now be applied to solve this stochastic optimization problem. For example, one can implement an exhaustive research if the dimensions/size of the area/environment where the sensor node evolves is known. In such case, the area is divided into a grid, then the likelihood function (cf. equation 4.30) is evaluated at each center of the squared grid in order to find the estimated position $(\hat{x}, \hat{y})$ giving the maximum value (cf. in [180], [181], [182]).

Note that the accuracy of the resulting position is related to the resolution of the chosen grid. In fact, it represents a trade off with the number of computations as a higher resolution means more values to evaluate and thus, it implies a possible long execution time before the maximum value is found. Instead, with fewer points to evaluate, a larger grid search implies a faster research, but with a chance to get a less accurate estimated position. To reduce such execution time, iterative algorithms for nonlinear optimization function are more appropriate in order to solve equation 4.30. Some of them have been proposed in the literature as : the Nelder-Mead simplex [163], the conjugate gradient descent [172] or the *Levenberg-Marquardt* [183].

RSSI-based indoor localization by maximum likelihood estimation approach After characterizing a bit how signal was attenuated in an indoor environment in Chapter 3 and defining a maximum likelihood estimator in the above section of this chapter 4, we implement a RSSI-based maximum likelihood localization approach (MLE) applied to a log-normal shadowing model to evaluate indoor localization accuracy.

We denote $\text{RSSI}_{j,\mathcal{L}}$ the j^{th} RSSI sample measured by the mobile node on packets coming from a given landmark $\mathcal{L}$. We write:

$$\text{RSSI}_{j,\mathcal{L}} \approx \text{PL}_{\mathcal{L}}(d_0) + 10.\eta_{\mathcal{L}} \log_{10} \left(\frac{d_{\mathcal{L}}}{d_0}\right) \tag{4.30}$$

where $d_{\mathcal{L}}$ is the distance to landmark $\mathcal{L}$ that we try to estimate, η a positive constant that

represents the environment path loss exponent and d_0 the reference distance where the value of the attenuation is equal to $\mathrm{PL}_{\mathcal{L}}(d_0)$. We have demonstrated in Chapter 3 that this model is not fully valid in a practical indoor context. To confirm quantitatively this assertion, we consider the problem of localization based on RSSI. We show, through experiments, that the location accuracy with the use of Maximum Likelihood estimation algorithm-based on RSSI, is slightly improved than the simple trilateration mechanism and even than a random localization.

More precisely the confidence interval stays in 1 meter. Rewrite equation 4.30 as:

$$\mathrm{RSSI}_{j,\mathcal{L}} = \mathrm{PL}_{\mathcal{L}}(d_0) + 10\eta_{\mathcal{L}} \log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right) + \mathcal{N}(0, \sigma^2_{\mathcal{L}})\,, \tag{4.31}$$

where $\mathrm{RSSI}_{j,\mathcal{L}}$ denotes the j-th attenuation observed in the unknown location (x, y) received from the landmark $\mathcal{L}$, $d_{\mathcal{L}}$ the distance between the landmark $\mathcal{L}$ and the location (x_{BN}, y_{BN}) of the blind mobile node. $\mathcal{N}(0, \sigma^2_{j,\mathcal{L}})$ is the noise associated to the landmark $\mathcal{L}$ for the j-th observation. $J_{\mathcal{L}}$ denotes the total number of observations from the landmark $\mathcal{L}$, and L the number of landmarks. For simplicity, let us write $\epsilon_{j,\mathcal{L}} = \mathcal{N}(0, \sigma^2_{j,\mathcal{L}})$. We assume that the random variables $\epsilon_{j,\mathcal{L}}$ are independent, Gaussian, with zero-mean and variances $\sigma^2_{\mathcal{L}}$ depending on the landmark.

Let us also write: $\alpha_{\mathcal{L},0} = \mathrm{PL}_{\mathcal{L}}(d_0)$ and $\alpha_{\mathcal{L},1} = \eta_{\mathcal{L}}$.

Here, as it is noticed in the previous section, the regression parameter $\alpha_{\mathcal{L}} = (\alpha_{\mathcal{L},0}, \alpha_{\mathcal{L},1})$ and the noise variance $\sigma^2_{\mathcal{L}}$ do depend on the landmark nodes through index $\mathcal{L}$.

The maximum likelihood estimator of the location (x, y) writes:

$$\begin{aligned}(\hat{x}, \hat{y}) \;=\; \arg \min_{x\in(0,X), y\in(0,Y)} & \sum_{\mathcal{L}=1}^{L} J_{\mathcal{L}} \log(\sigma^2_{\mathcal{L}}) + \\ & \sum_{\mathcal{L}=1}^{L} \sigma_{\mathcal{L}}^{-2} \sum_{j=1}^{J_{\mathcal{L}}} (A_{j,\mathcal{L}} - \alpha_{\mathcal{L},0} - \alpha_{\mathcal{L},1} \log_{10} d_{\mathcal{L}}(x, y))^2\,,\end{aligned} \tag{4.32}$$

where (x, y) are 2D coordinates of the blind mobile node named: (x_{BN}, y_{BN}). We have determined numerically the minimum, using a grid with step 50 cm and with $X = 3$ m and $Y = 4$ m.

The regression parameter estimation is based on the least square criterion. That writes

$$\hat{\alpha}_{\mathcal{L}} \;=\; (H^t_{\mathcal{L}} H_{\mathcal{L}})^{-1} H^t_{\mathcal{L}} A_{\mathcal{L}} \tag{4.33}$$

where $A_{\mathcal{L}}$ is the sequence of the attenuations observed from the landmark $\mathcal{L}$ and the design matrix

$$H_{\mathcal{L}} = \begin{bmatrix} 1 & 10\log_{10}(d_{\mathcal{L},1}) \\ \vdots & \vdots \\ 1 & 10\log_{10}(d_{\mathcal{L},K_{\mathcal{L}}}) \end{bmatrix}$$

where $K_{\mathcal{L}}$ is the number of the observations received from the landmark $\mathcal{L}$ in the *learning* data set. The estimated variance is given by

$$\hat{\sigma}^2_{\mathcal{L}} = \frac{(A_{\mathcal{L}} - H_{\mathcal{L}}\hat{\alpha}_{\mathcal{L}})^t (A_{\mathcal{L}} - H_{\mathcal{L}}\hat{\alpha}_{\mathcal{L}})}{(K_{\mathcal{L}} - 2)}$$

In the next section we present localization accuracy of Maximum Likelihood estimation (MLE) approach.

4.2.4 Comparison of RSSI-based maximum likelihood estimator and trilateration indoor localization results

Experimental protocol To evaluate the localization accuracy on the sole position estimation, we separate each experiment in two online steps as described in section 4.2.1. We use 20 positions as the *learning data set* to learn the α parameters, that is $\mathrm{PL}_{\mathcal{L}}(d_0)$ and $\eta_{\mathcal{L}}$ parameters w.r.t. the five landmarks $\mathcal{L}$. The rest of the positions, *i.e.* 28 positions, is the *testing data set* for testing the localization accuracy.

The results are reported in Figure 4.14. We observe that the distance error between the estimated location and the true location is a little better for Maximum Likelihood estimator (MLE). At first the mean of the error over all positions is 1 meter with MLE instead of 1.51 meters with simple trilateration and 1.84 meters with the random localization (RL). However the presence of large errors is also observed whereas 12 positions are estimated at less than 80 cm from the true location. We may conclude that the localization based on the model of attenuation given by equation 4.30 is not able to achieve an accuracy better than 1 meter. This result may be enhanced by taking into account the origin of large discrepancy of some locations. For that, some parameters should be considered : the height of the room, a more important number of landmarks and also the environmental location of the landmarks into the office.

Preliminary results of maximum likelihood and trilateration mechanism From our experimentation and results presented in Figure 4.14, we provide an affirmation of Elnahraw *et. al* [116] assumptions and additionally give information about parameters that should be determined. Due to the inability of the log-normal shadowing propagation model to capture other propagation effects than the distance in indoor environment, RSSI does not give much information about the distance even if the measurements and calculations it involves are less complicated than other localization metrics. In that context, the log-normal shadowing model is not well-suited in indoor environment for localization purpose due to high variability observed on the RSSI that can not be only explained by the log-normal shadowing model. Our studies highlight a certain dependency on material while the battery pack produces attenuation. That appears clearly when the MLE approach is used to localize : the results are very poor even if they are slightly better than those of trilateration and even of a random localization. This is mainly due to the presence of multi-path propagations, involving the effect of walls on the antenna gain, and/or occlusions.

In the following, we propose and evaluate a family of algorithms to enhance indoor positioning of wireless devices using RSSI measurements from data packets coming from multiple sources. We base our study on simulations and on experimental measurements campaigns in real environments.

Recall that, using one-hop localization algorithm in wireless sensor network (WSN) broadly supposes the presence of sensor nodes with initially unknown location information called blind mobile nodes. The location of each blind mobile node is estimated using inter-sensor measurements like distance and location information of nodes with known location (named landmarks or references nodes). Reference nodes positions can be obtained in a global coordinate system like GPS or manually when they are deployed at known coordinates.

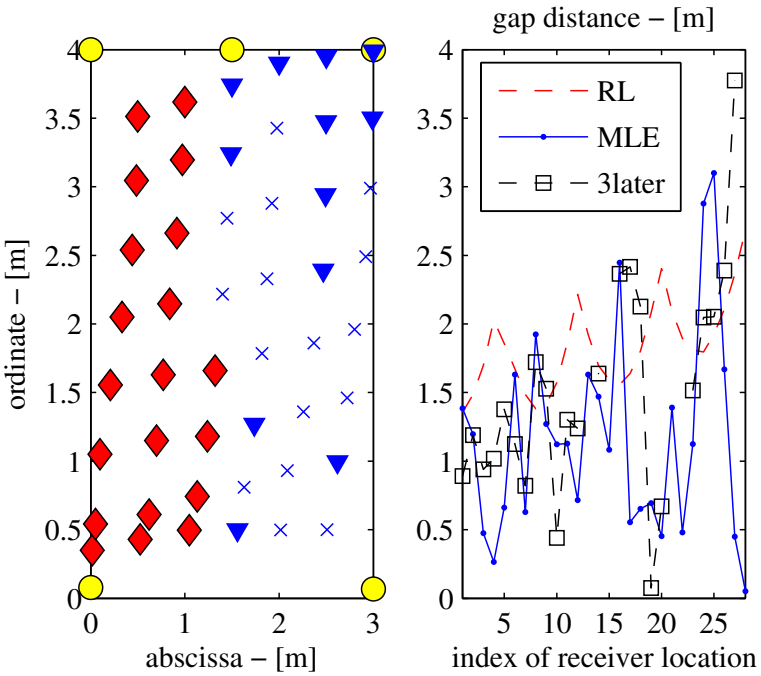


Figure 4.14: Left graph: landmarks location represented by yellow circles. Learning database from a target locations, represented by diamonds. *Testing* dataset locations indexed from 1 to 28, from top to bottom and from left to right. Triangle are location computed at less than 0.80 cm from the true position. Right graph: dashed line is the average gap distance from random localization. Solid line is the gap distance between the estimated location and the true location from MLE and square dashed line for trilateration (named "3later").

4.2.5 Localization System Overview

To evaluate and compare different localization strategies that will be presented in the following, we use the same localization setup/procedure as well as the same experimental protocol presented below. Recall, our proposed localization methods using the RSSI have been implemented and tested in real indoor environments. Our proposals performance have been evaluated in Chapter 3.

Localization procedure All sensors nodes in our testbeds, including landmarks and blind mobile nodes, are Tmote Sky motes, using IEEE 802.15.4 in the 2.4 GHz band to communicate. Nevertheless, the method we propose should be applicable at least to all types of narrowband RF transmissions in the same frequency band. This frequency band is open and devices operating in the same frequency band may impact the frame reception ratio as well as the reading RSSI. As such interferences can be expected in any real scenario, we recall we did not seek to conduct our experiments in isolated rooms.

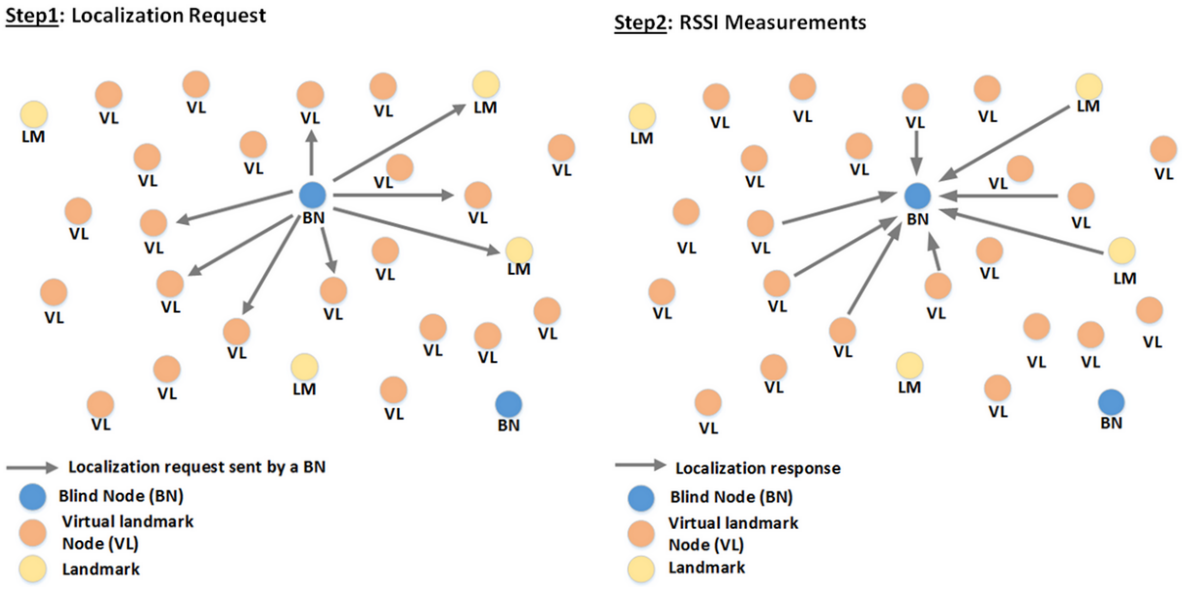
Landmarks as well as the blind mobile node are placed at a 1.25 m height from the ground, which is a standard height, for example corresponding to the hip of a human. In addition, two persons are moving inside each environment and impacting radio propagation. The mobile node is moved by an operator in several tested positions chosen randomly and different for each testbed.

RSSI values are used to determine the distances that separate a blind mobile node, willing to determine its position, and a set of landmarks also known as anchor nodes with known locations.

The collection of RSSI values require to design a communication protocol. Localization request is first sent by any blind mobile node to landmarks. The localization procedure is as follows as shown in Figure 4.16.

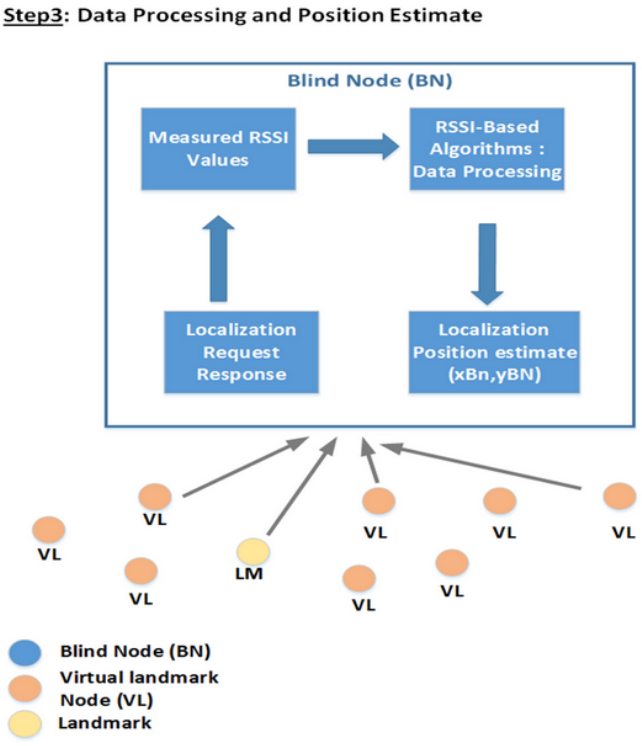
1. At each location, the mobile node tries to estimate its position by broadcasting a localization request that is received by all landmarks in its range.
2. Each landmark within the communication range of the requested blind mobile node then answers by broadcasting 100 beacons packets at a rate of one packet per second and then stops emitting until another request is received. In such communication structure, when landmarks receive a localization request from a blind node, they (landmarks) reply following a pseudo-random scheduling that is specific to each of them. Furthermore, depending on the desired configuration, parameters of sending packet frequency can be modified on each landmark, in order to send localization response to a blind node every millisecond or even nanosecond for example.
3. The blind mobile node collects position-related information including the RSSI values measured/sensed on a set of data packets sent by any landmark in its range. At each received message, a blind node measures an RSSI value that can be different from the previous one. Then, we use statistical techniques to improve the localization accuracy based on these RSSI measurements. Thus, no additional equipment is used, only the supplied Moteiv Tmote Sky nodes. Then, the blind node directly plugged to an Eee PC estimates online in the aftermath the propagation model parameters according to the deployment position of each landmark while using equations 4.5 and 4.7. Such information is used by the blind node (connected to an Eee PC) to compute its position from the data set. This process is repeated at all positions in each environment. Moreover, notice that Tmote Sky sensor nodes are different from commercial sensors (e.g. those from *Coronis* System) which have more powerful embedded processors, therefore such computations are possible on the sensor node itself.

Even if the energy consumption is not the target of this thesis, the communication protocols of our proposals are designed in the way that can be labeled to more respect the energy saving requirements of the wireless sensor network. Landmark in a confined/indoor area are in standby mode until a localization request message is received from any blind mobile node. Then a blind mobile node computes itself its position using any of the proposed indoor localization algorithms. This mechanism saves probably more energy than letting landmarks to continuously send localization information. Indeed, for a given packet received by a blind mobile node, the number of bits used by our communication protocols takes into account the value of the measured RSSI value coded under 8 bits, the landmark's position and its address that are between 4 and 20 bytes; as well as the sequence number of the packet, one byte. Such combined information makes a total number of roughly 8 bytes which is much lower than the required throughput by the IEEE 802.15.4 standard (i.e. 8 KBps). Thus, 1000 packet per second can be received on a blind node without contention. In addition, regarding the related work on the energy consumption, including those from Heinzelman *et al.* [184] [185], in transmission mode of packets, the energy consumption of the network depends on the following factors: the number of sent bits, i.e. the packet size, the square of the radius or the communication distance as well as two other constants. The first factor depends on the electronic and the second one is the amplification factor which increases with the distance. Note that that we communicate over short distances in indoor environments and the transmission range of the Tmote Sky in indoor is 50 m. The combination of such elements result in less energy consumption, as even if the packet sending



(a) localization step1

(b) localization step2



(c) localization step3

Figure 4.15: Our localization system overview

Figure 4.16: An example of view of localization step in a testbed

frequency is increased at about 1 packet per millisecond, our algorithms meet the required technical specifications of the IEEE 802.15.4 in terms of maximum throughput.

The landmarks positions and the successive mobile node positions are reported in Figures 3.2, 3.3 and 3.5 in Chapter 3. Landmarks are randomly deployed and their number is different for each tested environment.

Moreover, other works [86, 175] realized in realistic indoor testbeds using localization with sensor networks based on min-max algorithms and landmarks selection have demonstrated that better landmarks nodes with low standard error are situated in the extremity of the building, not in the corridors. According to the authors, the corner landmarks nodes are located in positions less sensitive to multipath fading and are considered as more informative ones.

Below, we describe the communication protocols and structure that respect all proposed localization algorithms.

Localization communication protocols When using one-hop localization algorithms, only RSSI values coming from the landmarks are exploited by any blind mobile node. In ideal propagation condition cases without or less packet losses, the number of RSSI values that can be collected by a blind mobile node in 1 min is shown in Table 4.2.

Table 4.2: The collected number of RSSI values in our communication structure for one-hop localization algorithms.

	Number of RSSI values in a minute
Blind mobile node to localize ⇐ landmark node	140

However, when dealing with distributed localization algorithms, that is a blind mobile node can estimate its position and those of its neighboring sensor nodes in the network, the communication structure changes. This new communication structure integrates the concept of virtual landmarks. The virtual landmarks are landmarks with unknown positions, in which only the measured RSSI values from data packets they send are involved in the blind mobile node localization process. Note that from its own point of view, each independent sensor node with unknown position is a blind node and the rest of the nodes in the network are virtual landmarks

and landmarks. Such consideration is valid for each node with unknown position. Then, to localize itself in a sensor network, a blind node sends a localization request. This request is received by all landmarks in its range and in the same time by virtual landmarks. As shown in Table 4.2 a blind mobile node gathers RSSI data measured from data packets send by all landmarks in its range as well as all neighboring nodes with unknown positions called virtual landmarks. Thus a blind node is a virtual landmark from the other blind nodes point of view. Using our proposed distributed localization algorithms, a blind node is able to localize itself while estimating virtual landmarks (*i.e.* other blind nodes) positions. In this way, any blind mobile node as well as any virtual node can have the complete map of the network with respective positions. The position accuracy of these three algorithms are evaluated in three different indoor environments represented in Figures 3.2, 3.3 and 3.5 in Chapter 3 and also illustrated in this chapter (cf. Figures 4.20(a), 4.20(b) and 4.20(c)). These scenarios characteristics are reported in Table 4.4, including the number of people in each testbed, their movements and objects obstructions.

4.3 Bias-based methods

4.3.1 Excluding a single aberrant landmark

Uncontrolled scenario from our measurement campaigns show that if some sources are coherent, some are clearly incoherent with the others. This phenomenon can be explained by the radio propagation (e.g. multipath), or by poor hardware and excluding these incoherent data sources from the dataset should lead to a better accuracy. We propose a dynamic method to identify abnormal landmarks and to reduce the effects of corrupted RSSI values measured from data sent by such landmarks, and finally replacing these RSSI measurement by a constant bias.

4.3.1.1 Biased Log-Normal Shadowing Model (B-LNM)

Due to practical situation encountered from our testbeds presented in Chapter 3, we notice some variations on the standard deviation even if the environment and the propagation condition remain the same. Hence, the idea to consider the LNSM parameters are based on landmark locations.

This section introduces a method that potentially reduces the effects of a single outlier landmark from the set of landmarks used by the mobile node for localization if the mobile node considers that the associated measurements do not improve localization accuracy. Furthermore, as shown in Chapter 3, measurements coming from landmarks can present a high standard deviation, or a distribution that does not fit a normal distribution (cf. histogram of RSSI values in Chapter 3). In this case, values coming from this landmark should not be included in our estimation, as they will only decrease accuracy. In the first method presented here, we compensate this effect by introducing a constant bias, that constitutes an additional variable to estimate, to replace the log-normal shadowing model of the measurements associated to this landmark. We select the aberrant landmark by comparing the global likelihood values obtained when each landmark is considered as outlier. This method requires to use RSSI measurements from at least 4 landmarks for the algorithm to work. In the case there is abnormal landmarks, RSSI measurements from a minimum of three landmarks are used to compute the blind node 2D coordinates. Therefore, in each of the experiment, we estimated the distances to the third closest landmark for each blind mobile node position.

Table 4.3: Main notations

$PL(d)$	Path loss over distance d
$PL_{\mathcal{L}}(d)$	Path loss between mobile node and a landmark $\mathcal{L}$
η	Attenuation factor
$\eta_{\mathcal{L}}$	Attenuation factor between mobile node and a landmark $\mathcal{L}$
d_0	Reference distance
$d_{\mathcal{L}}$	Distance between mobile node and a landmark $\mathcal{L}$
σ	Standard deviation of the channel noise
$\sigma_{\mathcal{L}}$	Standard deviation of the channel noise between mobile node and a landmark $\mathcal{L}$
β	Unknown constant bias
K	number of landmarks
$N_{\mathcal{L}}$	Number of RSSI observations from a landmark $\mathcal{L}$ to mobile node

A summary of the notations used in the following paragraphs is reported in Table 4.3. Let us denote by $\mathrm{RSSI}_{j,\mathcal{L}}$ the j^{th} RSSI sample measured by the mobile node on packets coming from a given landmark $\mathcal{L}$. If we denote by $\mathrm{PL}_{\mathcal{L}}(d_0)$, $\eta_{\mathcal{L}}$ and $\sigma_{\mathcal{L}}$ the log-normal shadowing model specific parameters for landmark $\mathcal{L}$, we can rewrite the equation (3.2) in the general case, replacing measurements coming from a given landmark $\mathcal{O}$ by a constant bias, β, as follows:

$$\mathrm{RSSI}_{j,\mathcal{L}} = \left(\mathrm{PL}_{\mathcal{L}}(d_0) + 10.\eta_{\mathcal{L}} \log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right)\right).\mathbb{1}_{\mathcal{L}\neq\mathcal{O}} + \beta.\mathbb{1}_{\mathcal{L}=\mathcal{O}} + \mathcal{N}(0,\sigma_{\mathcal{L}}^2)\,, \tag{4.34}$$

where $d_{\mathcal{L}}$ is the distance to landmark $\mathcal{L}$ that we try to estimate and $\mathbb{1}$ is the indicator function (equal to 1 when the subscript expression is true, to 0 otherwise).

4.3.1.2 Biased-Maximum Likelihood Location Estimation Method (B-MLE): selecting the outlier landmark

Combining all the measured values altogether, we can apply a maximum likelihood estimator on this new model to compute the likelihood expressions in the case where landmark $\mathcal{O}$ is considered as abnormal. If we denote by $N_{\mathcal{L}}$ the number of samples received from landmark $\mathcal{L}$, the likelihood function is expressed as follows for every landmark $\mathcal{L} \neq \mathcal{O}$:

$$L_{\mathcal{L}}(x,y) = -N_{\mathcal{L}}.\log\sigma_{\mathcal{L}}^2 - \sum_{j=1}^{N_{\mathcal{L}}}\left(\frac{\mathrm{RSSI}_{j,\mathcal{L}} - \mathrm{PL}_{\mathcal{L}}(d_0) - 10\eta_{\mathcal{L}}\log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right)}{\sigma_{\mathcal{L}}}\right)^2. \tag{4.35}$$

And for the outlier landmark, $\mathcal{O}$, it becomes:

$$L_{\mathcal{O}}(\beta) = -N_{\mathcal{O}}\log\sigma_{\mathcal{O}}^2 - \sum_{j=1}^{N_{\mathcal{O}}}\left(\frac{\mathrm{RSSI}_{j,\mathcal{O}} - \beta}{\sigma_{\mathcal{O}}}\right)^2. \tag{4.36}$$

The global likelihood function from the data set, reflecting the coherence of the whole system when landmark $\mathcal{O}$ is the outlier one is then simply the sum of the previous expressions over the

K landmarks:

$$L_{\mathcal{O}}(x, y, \beta) = \sum_{\mathcal{L}=1;\mathcal{L}\neq\mathcal{O}}^{K} L_{\mathcal{L}}(x, y) + L_{\mathcal{O}}(\beta) \tag{4.37}$$

The global likelihood function without abnormal landmark is also expressed as:

$$L_0(x, y) = \sum_{\mathcal{L}=1}^{K} L_{\mathcal{L}}(x, y)$$

Equation 4.37 can be written for each of the K landmarks. This defines, for each of the K situations, a maximum likelihood estimation problem that can be solved to deduce the corresponding position and bias:

$$(x_{\mathcal{O}}, y_{\mathcal{O}}, \beta_{\mathcal{O}}) = \arg\max_{x,y,\beta}\{L_{\mathcal{O}}(x, y, \beta)\} \tag{4.38}$$

The maximum likelihood estimation for the model without bias is written as:

$$(x_0, y_0) = \arg\max_{x,y}\{L_0(x, y)\} \tag{4.39}$$

In the expression of equation 4.37, the $L_{\mathcal{L}}(x, y)$ terms does not depend on β while the $L_{\mathcal{O}}(\beta)$ term does not depend on x or y. Therefore the maximization of equation 4.38 can be achieved by separating the variables. In particular, the bias expression that maximizes $L_{\mathcal{O}}(\beta)$ is straightforward:

$$\beta_{\mathcal{L}} = \frac{1}{N_{\mathcal{L}}} \sum_{j=1}^{N_{\mathcal{L}}} RSSI_{j,\mathcal{L}} \tag{4.40}$$

If we also consider the case in which no landmark is considered abnormal, we get $K+1$ likelihood values to compare. We then have to select the situation that yields to the best (*i.e.* highest) likelihood value to identify the most aberrant landmark.

Results, detailed in section 4.3.3 show that sometimes the best likelihood corresponds to the situation in which no landmark is outlier. It is most likely that in some situations, accuracy can be further improved by considering more than one landmark as abnormal. The global flowchart of the algorithm is represented in Figure 4.17.

4.3.2 Excluding multiple landmarks: Extended Biased-Maximum Likelihood Location Estimation Method (xB-MLE)

As mentioned above, there are cases in which it would be profitable to consider more than one outlier landmarks. To study the pertinence of such an Extended Biased-Maximum Likelihood Estimation (xB-MLE), we first examine the achieved accuracy when two landmarks are abnormals, then three landmarks in our various experimental setups. When more than one landmark is abnormal, each set of measurements is replaced by a distinct bias. This allows to keep the

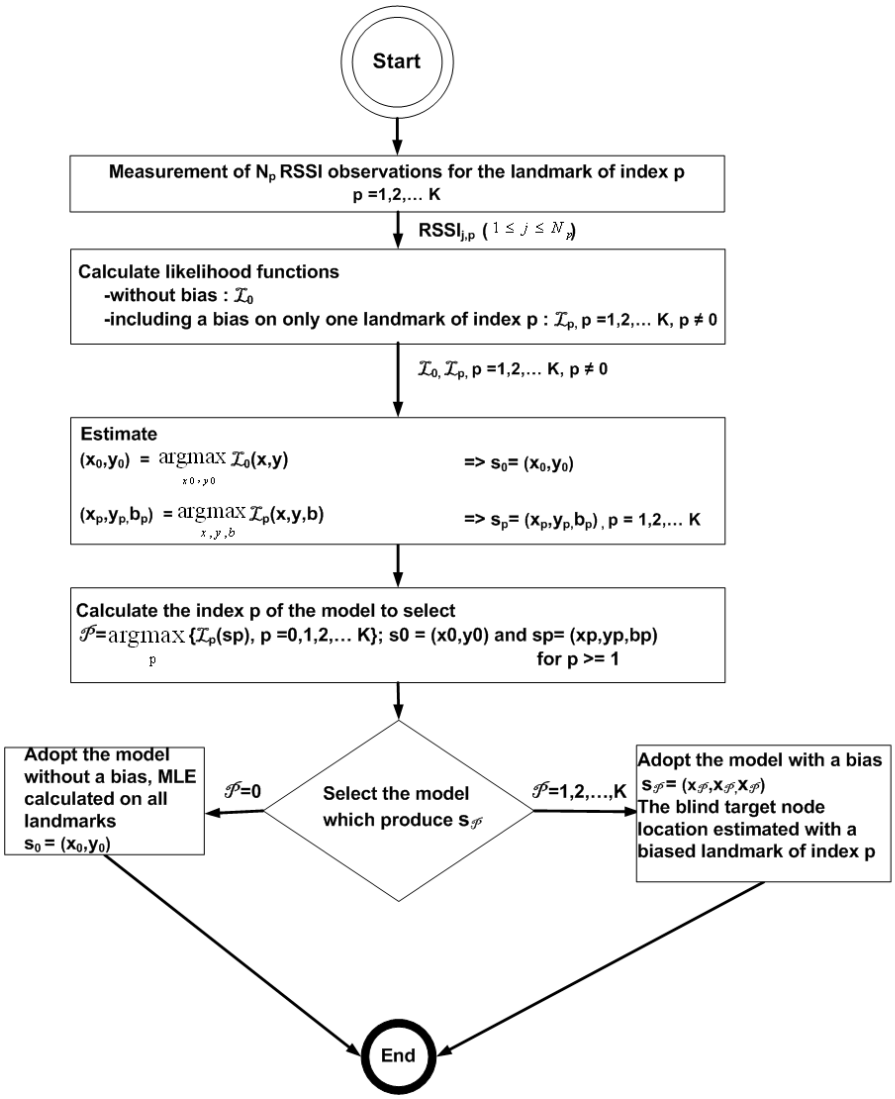


Figure 4.17: Flowchart of B-MLE algorithm

same maximization method as above, separating variables for the optimization. For example, in the case we consider two landmarks, $\mathcal{O}_1$ and $\mathcal{O}_2$, for exclusion, the global likelihood function is expressed as:

$$L_{\mathcal{O}}(x, y, \beta_1, \beta_2) = \sum_{\mathcal{L}=1;\mathcal{L}\neq\mathcal{O}_1;\mathcal{L}\neq\mathcal{O}_2}^{K} L_{\mathcal{L}}(x, y) + L_{\mathcal{O}_1}(\beta_1) + L_{\mathcal{O}_2}(\beta_2) \quad (4.41)$$

Nevertheless, this process does not scale. If considering one landmark as abnormal requires $K+1$ maximization operations, two outlier landmarks requires $1 + K + \binom{n}{2}$ maximizations. In order to reduce this global complexity, we derive a heuristic based on the computation of a bias gap, defined as the relative difference between the bias and the estimated average path loss. This heuristic is the corrected extension of B-MLE named B-MLEC (Biased-Maximum Likelihood Location Estimation extension Correction) and expressed as follows:

$$\mathcal{B}_{\mathcal{L}} = \left| \frac{\beta_{\mathcal{L}} - \left(\mathrm{PL}_{\mathcal{L}}(d_0) + 10\eta_{\mathcal{L}} \log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right)\right)}{\mathrm{PL}_{\mathcal{L}}(d_0) + 10\eta_{\mathcal{L}} \log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right)} \right| \quad (4.42)$$

This value is computed based on the bias expression of equation 4.40 and hence induces a relatively low computational cost. We then compare this value with a threshold, γ, and when $\mathcal{B}_{\mathcal{L}} \geq \gamma$, landmark $\mathcal{L}$ is considered as abnormal. In the same way if $\mathcal{B}_{\mathcal{L}} \leq \gamma$, the landmark of index $\mathcal{L}$ is considered as normal. In the end, we only have one maximization operation to perform, as we can ignore all landmarks that do not satisfy this condition. For the present paper scenarios we empirically chose $\gamma = 0.05$, as this value corresponds to a generally admitted rejection threshold. To confirm this threshold, we evaluated the performance of our algorithms with different threshold values (ranging from 0.004 to 0.7). Figure 4.18 represents the average localization error achieved by B-MLEC in our three testbeds and shows that error increases from a threshold value of 0.1. The best value varies with the testbed, but 0.05 seems a fair choice.

The flow chart of this algorithm can be found in Figure 4.19.

4.3.3 Experimental results

4.3.3.1 Experimental setup

The position accuracy of these three algorithms was evaluated in three different indoor environments, represented in Figures 4.20(a), 4.20(b) and 4.20(c). These scenarios characteristics are reported in Table 4.4. In all scenarios, Wi-Fi access points are present and active in the room vicinity.

Two testbeds (named #2 and #3) are furnished classroom of different sizes. The other testbed (named #1) is a lightly furnished office. Testbed #3, measuring 8.77 m by 6.46 m, is part of a classroom that is divided in two parts. We only used one part for our measurements. 8 landmarks compose the testbed #3 while 5 for the testbed #1 and 4 for the last one (testbed #2). The number of landmarks in each testbed are chosen depending on the available hardware. In each environment, the receiving mobile node is positioned in different tested locations which are respectively 57 for the testbed #3, 48 for the testbed #1 and 47 for the testbed #2.

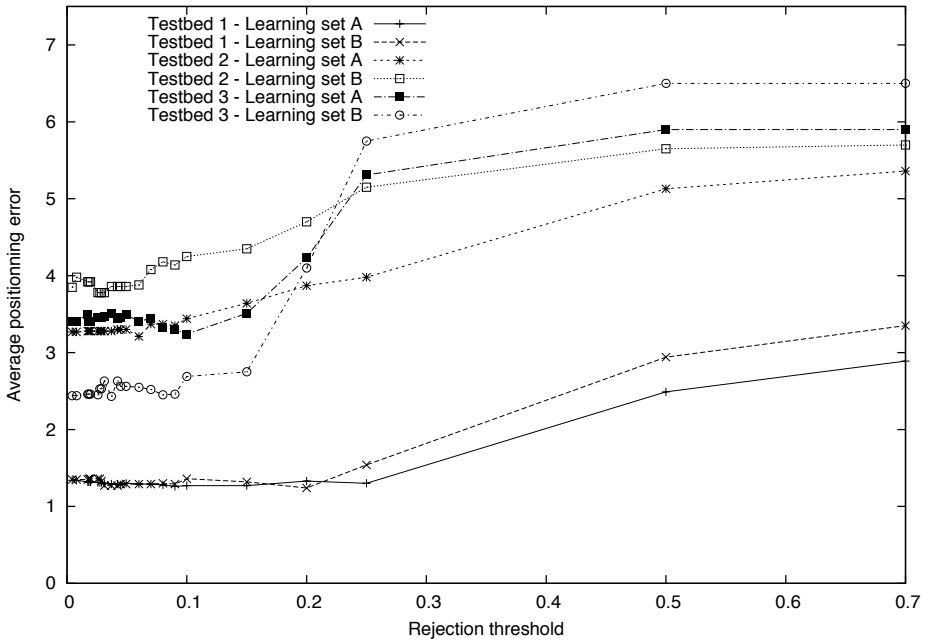


Figure 4.18: Average localization errors achieved by B-MLEC for different values of the rejection threshold in the various testbeds

Table 4.4: Testbeds characteristics

	Testbed 1 Fig. 4.20(a)	Testbed 2 Fig. 4.20(b)	Testbed 3 Fig. 4.20(c)
Room Size	4 m × 3 m	7 m × 7 m	8.77 m × 6.46 m
Moving people	2	2	Variable (open classroom)
Furniture, walls	woods, glass-fronted, metal		
Node position	Random grid		
Number of mobile node locations	48	47	57
Number of landmarks	5	4	8
Nodes height	1.25 m		
Emitted power	0 dBm		
MAC protocol	XMAC		
Operating system	Contiki		
Type of nodes	Moteiv TMote Sky		

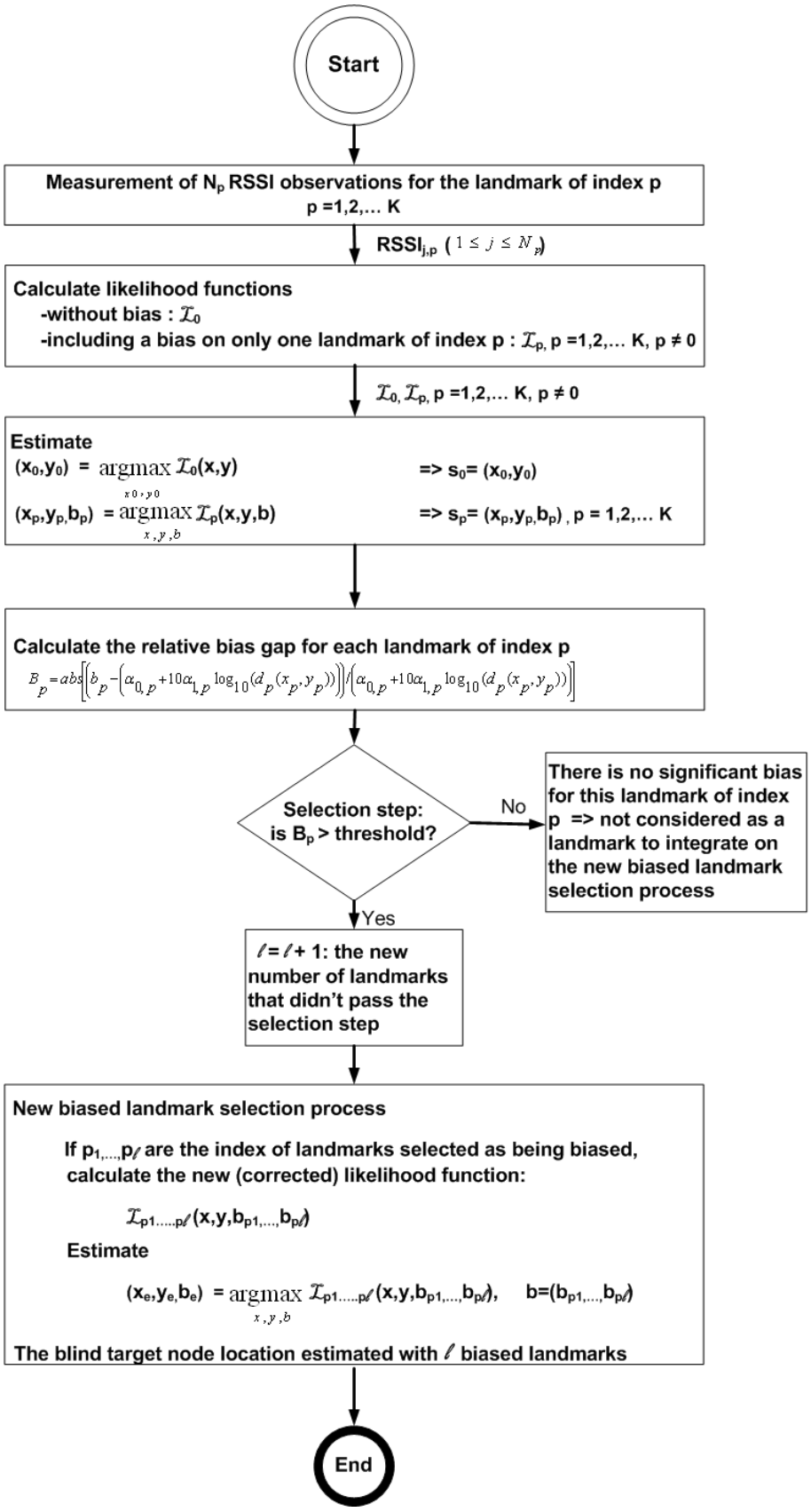


Figure 4.19: Flowchart of the extended B-MLE algorithm

We recall that each experiment is composed of two steps, which are both performed online. In the first step, we only estimate the values of the propagation model parameters ($PL(d_0)$, η and σ) for each landmark using least squares estimation. We use, for this first phase, a set of measurements coming from *learning data set* defined in section 4.2.1 and represented in Figure 4.20. As stated before, this estimation is done by the mobile and is not the results of a pre-calibration by an operator. Then, on the remaining position, we consider these parameters as constant and compare the three algorithms accuracy. This second set of positions used to compare localization algorithms constitutes the *testing data set*. Figure 4.20 divides the successive positions of the mobile node locations into three groups to identify the datasets we used for learning. The positions materialized by orange squares compose the smallest learning set (10 positions). The larger learning set (25 positions) is composed by the points identified by both the orange squares and the magenta triangles; it fully includes the smaller one. The blue stars are always part of the testing data set.

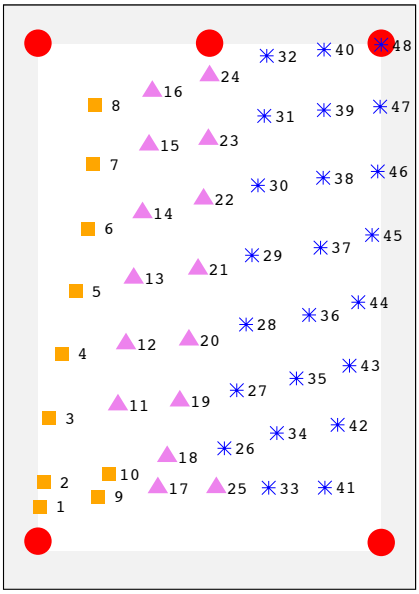
Main results are summarized by the 6 box-plots presented in Figures 4.21 and 4.22 and in Tables 4.5, 4.6 and 4.7. These graphs compare the different algorithms with a classical MLE estimator and with the random localization process to evaluate the contribution of RSSI. The results for B-MLE approach are obtained from all three testbeds, while the results when ruling out two landmarks, named 2B-MLE (2 landmarks biased maximum likelihood estimator), are only presented in two testbeds (a smaller (cf. testbed #1) and a bigger one (cf. testbed #3)). As we need at least three landmarks to localize, and based on the number of landmarks in each testbed, we will be able to present the results when ruling out three landmarks (named 3B-MLE for 3 landmarks biased maximum likelihood estimator) for only one testbed (testbed #3). Figures 4.21 and 4.22 represent the statistical properties of the achieved localization errors. They are analyzed under different angles and detailed in the subsequent paragraphs.

Table 4.5: Compared performance of the different algorithms in testbed #1

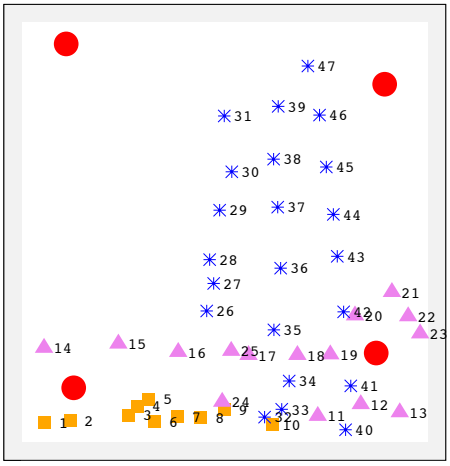
	Database	mean error	Median	STD
Random	A	1.79 m	1.74 m	0.31
MLE	A	1.29 m	1.20 m	1.29
B-MLE	A	1.39 m	1.29 m	0.85
2B-MLE	A	1.40 m	1.33 m	0.74
B-MLEC	A	1.30 m	1.20 m	0.85
Random	B	1.87 m	1.80 m	0.33
MLE	B	1.36 m	1.24 m	1.20
B-MLE	B	1.34 m	1.20 m	0.92
2B-MLE	B	1.39 m	1.41 m	0.75
B-MLEC	B	1.29 m	1.06 m	0.93

4.3.4 Results when excluding one landmark

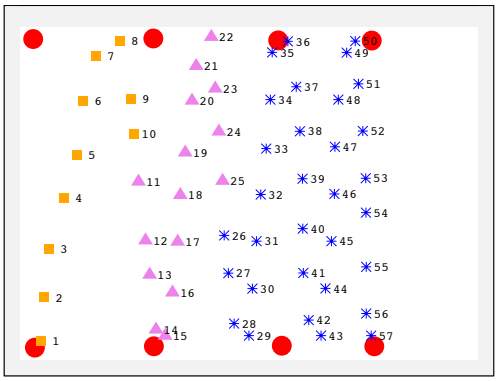
By comparing the results on the simple B-MLE on the different testbeds, we can see that B-MLE slightly improves accuracy over the classical unbiased MLE. Figure 4.24 represents the square error achieved by MLE and B-MLE with different learning sets sizes on the three testbeds. Each column represents one measurement point and the higher the points in the column are, the less accurate the method is. We can see that regardless of the testbed, the relative performance of the three algorithms is far from constant. Sometimes, excluding a landmark that appears aberrant really improves the positioning while sometimes (less often, though), it ends degraded.



(a) Testbed 1: (5 landmarks, 48 positions)

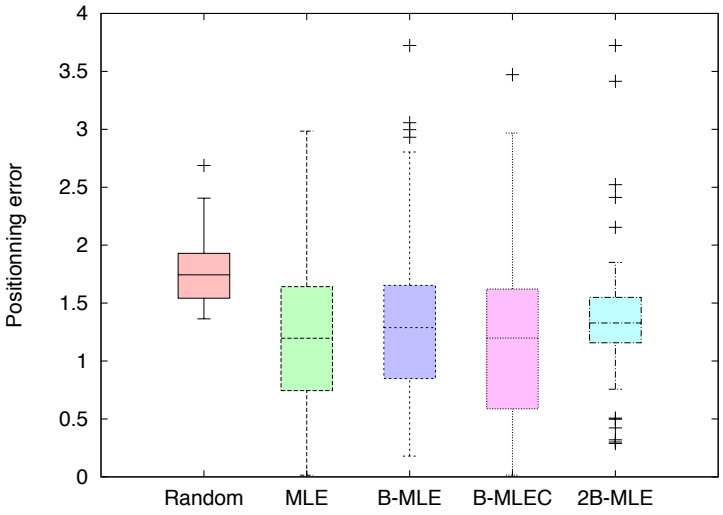


(b) Testbed 2: (4 landmarks, 47 positions)

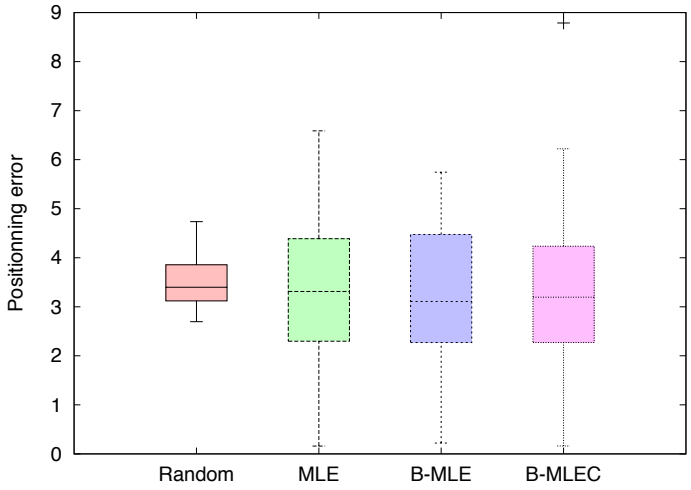


(c) Testbed 3: (8 landmarks, 57 positions)

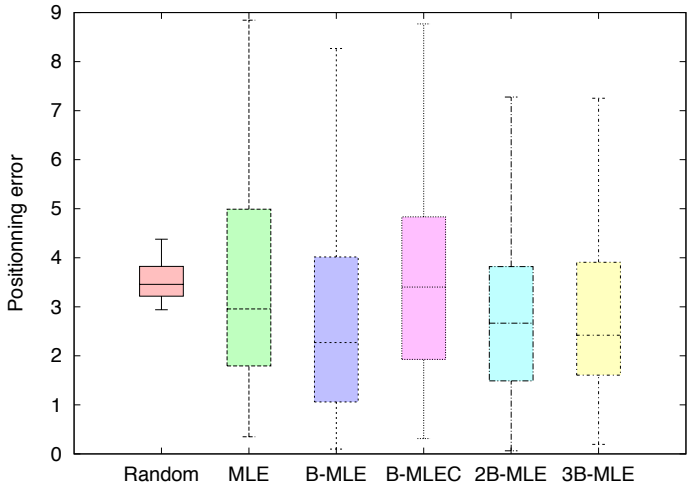
Figure 4.20: Locations of the landmarks (red dots) and identification of the positions composing the small learning set (orange squares) and the large learning set (orange squares and magenta triangles) in the three testbeds



(a) Testbed 1 - leaning set A (10 positions)

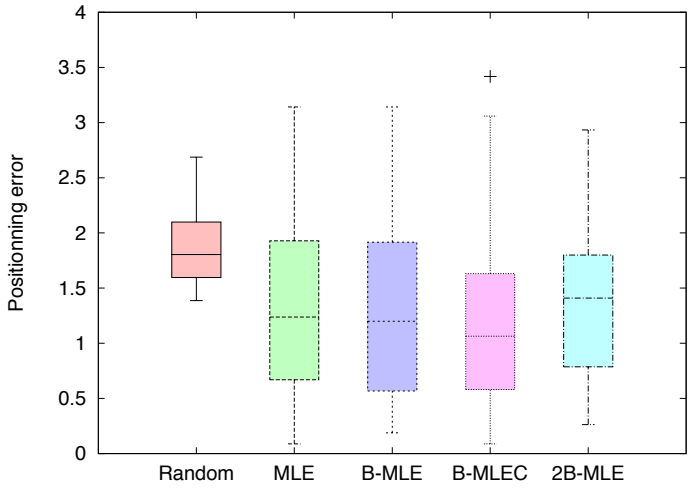


(b) Testbed 2 - leaning set A (10 positions)

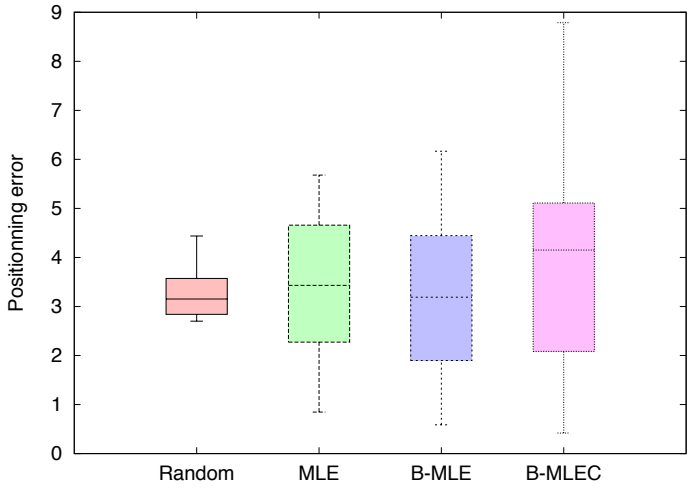


(c) Testbed 3 - leaning set A (10 positions)

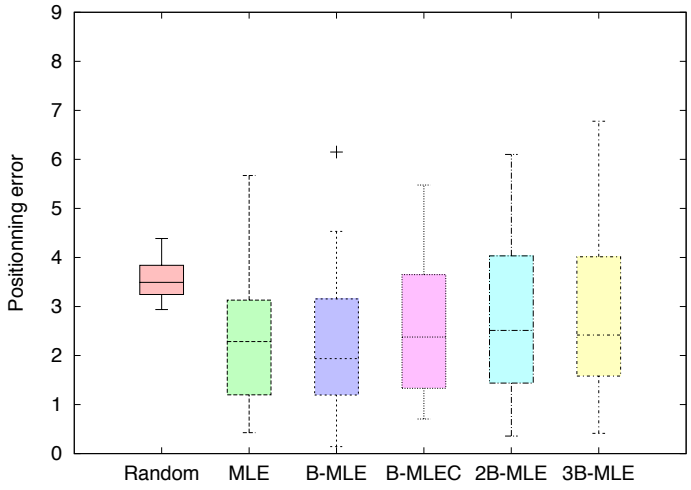
Figure 4.21: Box-plots of the localization errors achieved by the different algorithms in the various testbeds - Learning Set A (10 positions)



(a) Testbed 1 - leaning set B (25 positions)



(b) Testbed 2 - leaning set B (25 positions)



(c) Testbed 3 - leaning set B (25 positions)

Figure 4.22: Box-plots of the localization errors achieved by the different algorithms in the various testbeds - Learning Set B (25 positions)

Table 4.6: Compared performance of the different algorithms in testbed #2

	Database	mean error	Median	STD
Random	A	3.48 m	3.40 m	0.54
MLE	A	3.27 m	3.31 m	3.11
B-MLE	A	3.16 m	3.11 m	1.51
B-MLEC	A	3.30 m	3.20 m	1.76
Random	B	3.27 m	3.15 m	0.48
MLE	B	3.30 m	3.43 m	3.19
B-MLE	B	3.22 m	3.19 m	1.76
B-MLEC	B	3.86 m	4.01 m	2.38

Table 4.7: Compared performance of the different algorithms in testbed #3

	Database	mean error	Median	STD
Random	A	3.53 m	3.46 m	0.39
MLE	A	3.41 m	2.96 m	2.27
B-MLE	A	2.73 m	2.27 m	1.93
2B-MLE	A	2.72 m	2.67 m	1.59
3B-MLE	A	2.72 m	2.42 m	1.68
B-MLEC	A	3.49 m	3.40 m	2.05
Random	B	3.55 m	3.49 m	0.40
MLE	B	2.38 m	2.28 m	1.94
B-MLE	B	2.20 m	1.94 m	1.40
2B-MLE	B	2.73 m	2.51 m	1.49
3B-MLE	B	2.72 m	2.42 m	1.43
B-MLEC	B	2.56 m	2.38 m	1.40

Figure 4.23 represents the error mapped on the 2D map of testbed #1. This graph shows that there does not seem to be an environment-related reason to this, though, as the relative performance neither depends on the proximity to walls or structures, nor to the closeness to landmarks. It also shows that the points on which MLE fails are not the same as the points on which B-MLE exhibits a limited performance. Looking back at the global results, we can see that the overall performance is a bit better in testbed #3, which is the large classroom with 8 landmarks. This suggests that increasing the number of landmarks makes it more effective to remove one, but it is not systematic as confirmed by Figure 4.24(c).

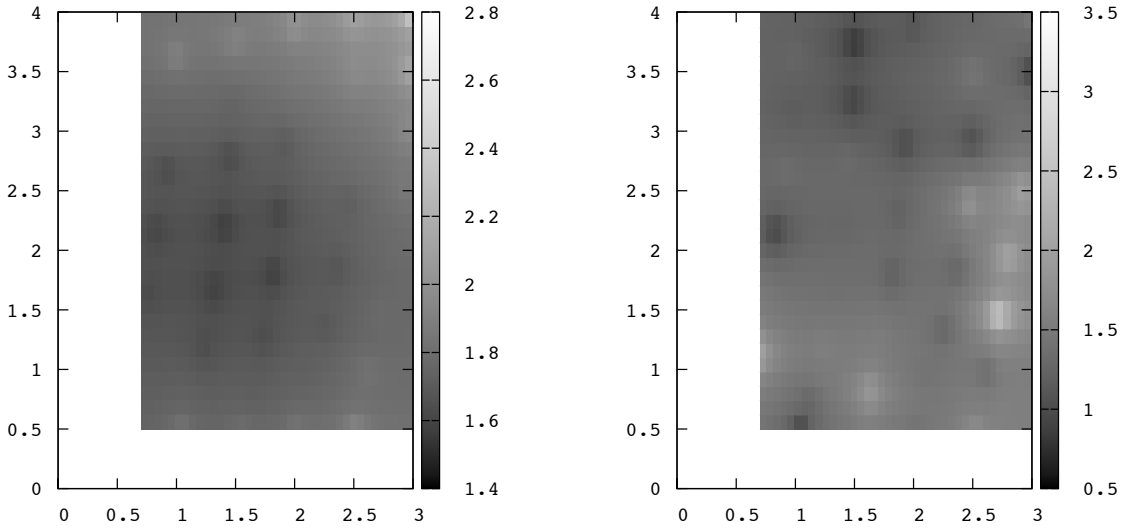
Similar conclusions hold concerning the size of the learning set. If increasing the number of samples often improves accuracy, it also often yields the same performance and sometimes degrades the situation. However, in the smallest testbed (#1), there are several outliers with the smallest learning set that disappear when the learning set size increases. This shows that the learning set size matters at least when the size of the room is small. Figures 4.23 and 4.24(a) show, however, that the locations at which B-MLE performs poorly are often the same with the two datasets. This shows that, if the decision taken by the algorithm regarding the abnormal landmark is similar in both cases, the performance improvement comes from a better estimation of the model parameters.

4.3.5 Results when excluding more than one landmark (xB-MLE)

As seen above, considering one abnormal landmark is not effective in all situations. In this section we are interested in what happens when we iterate this process, excluding successively two or three landmarks. The results presented in Figure 4.25 represent the relative performance of the different strategies on testbed #3. Coupled with the global results (see Figs. 4.21 and 4.22), we can draw the same conclusion: the strategy may be effective for some positions, but does not always have a positive effect. We can see, for example, that on the smallest testbeds (#1), 2B-MLE, which consists in reducing the effects of corrupted RSSI values from two landmarks, does not yield any improvement. Increasing the size of the learning data set only improves lightly the accuracy, reducing the number of outliers. On testbed #3, 2B-MLE slightly improves accuracy for the smallest learning set, but not with the largest. This result shows that it can be profitable to consider more than one landmark as abnormal and that the estimation of the propagation parameters can be performed more rapidly (*i.e.* a smaller learning database is required) when the number of sources is reduced.

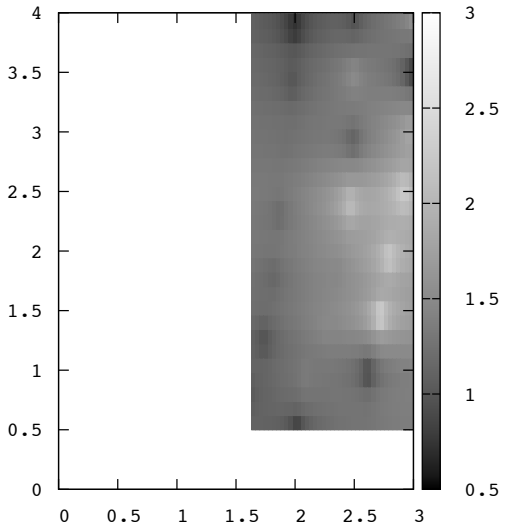
When looking at the threshold-based exclusion (B-MLEC described in section 4.3.2), we can notice different results on the three testbeds. On testbed #1, we can notice that the threshold-based approach is comparable to unbiased MLE and only slightly improves performance compared to B-MLE for smaller *learning data set*. However, for bigger *learning data set*, this threshold-based approach improves performance, as the standard deviation is reduced compared to MLE as well as to B-MLE. On this testbed #3, we can see that the adaptive selection of abnormal landmarks does not improve accuracy.

These results are confirmed by Figure 4.26, which represents, for each testbed, the percentage of positions among the testing positions for which each strategy is the most or the less efficient. If B-MLE and B-MLEC seem the most promising in the smallest testbed, the threshold does not seem appropriate in the large classroom (testbed 3). Surprisingly, the random strategy is often effective in the second testbed, followed by B-MLE and the adaptive strategy. The good performance of 3B-MLE on the large testbed is also encouraging.



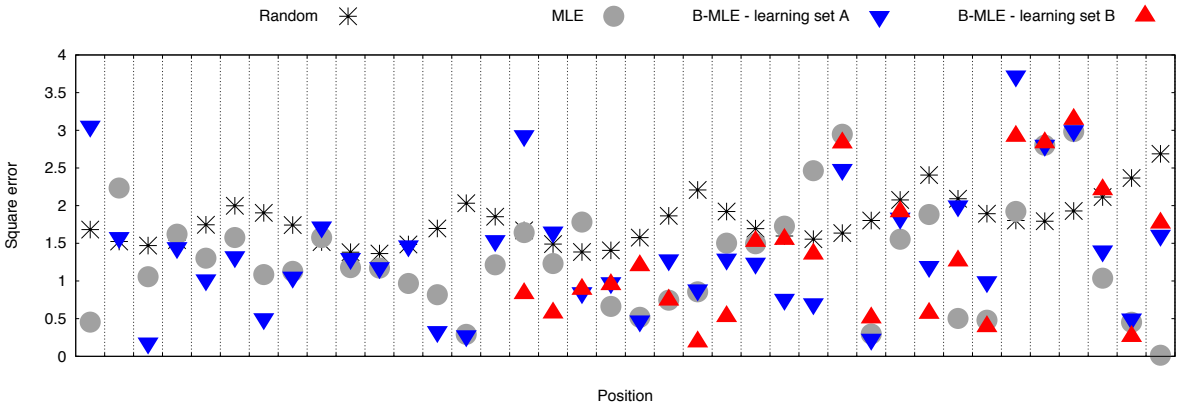
(a) MLE

(b) B-MLE ; leaning set A

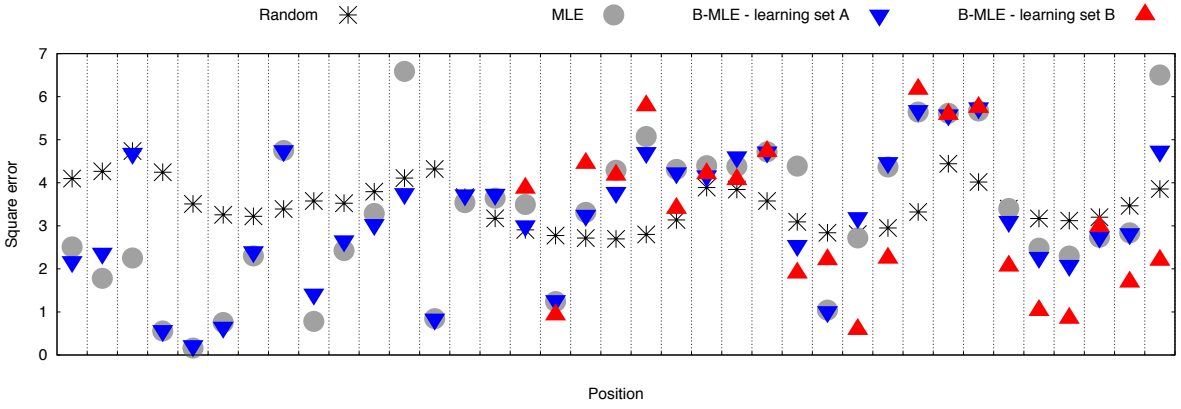


(c) B-MLE ; leaning set B

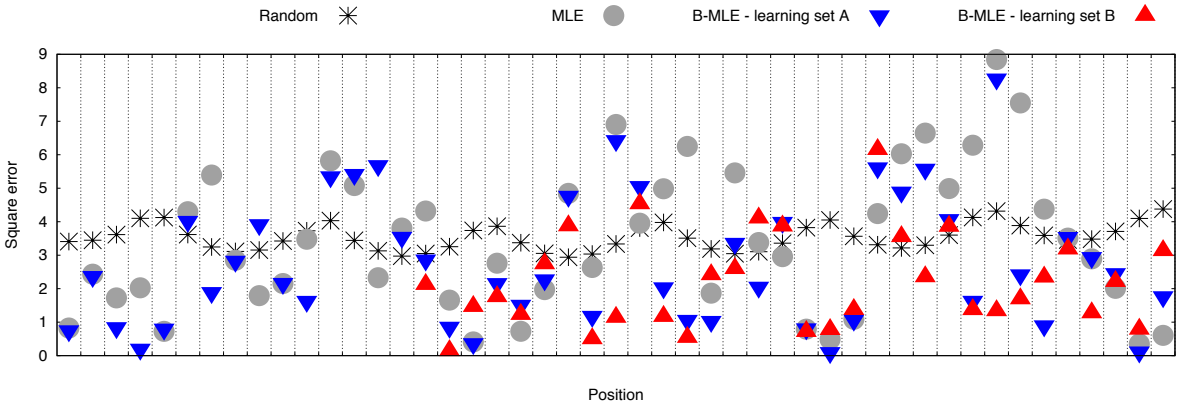
Figure 4.23: Positioning Error projected on testbed #1 map



(a) Testbed #1

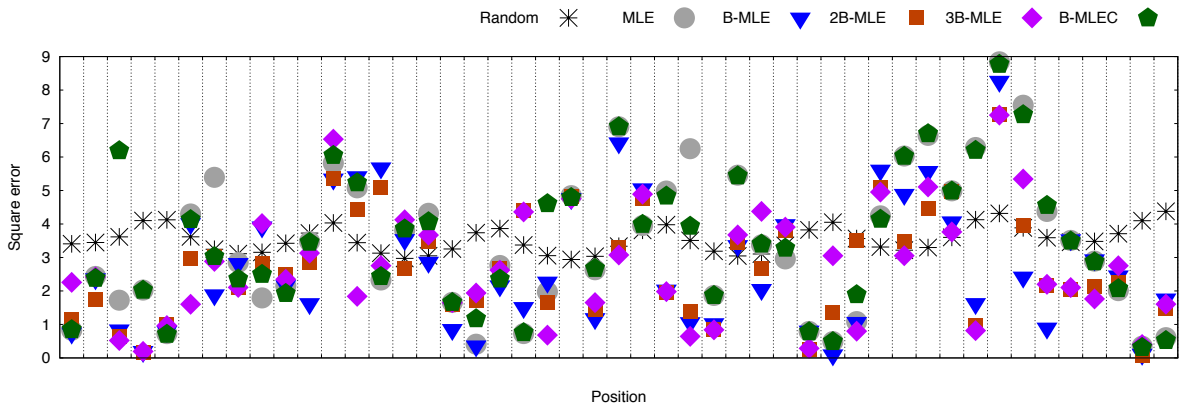


(b) Testbed #2

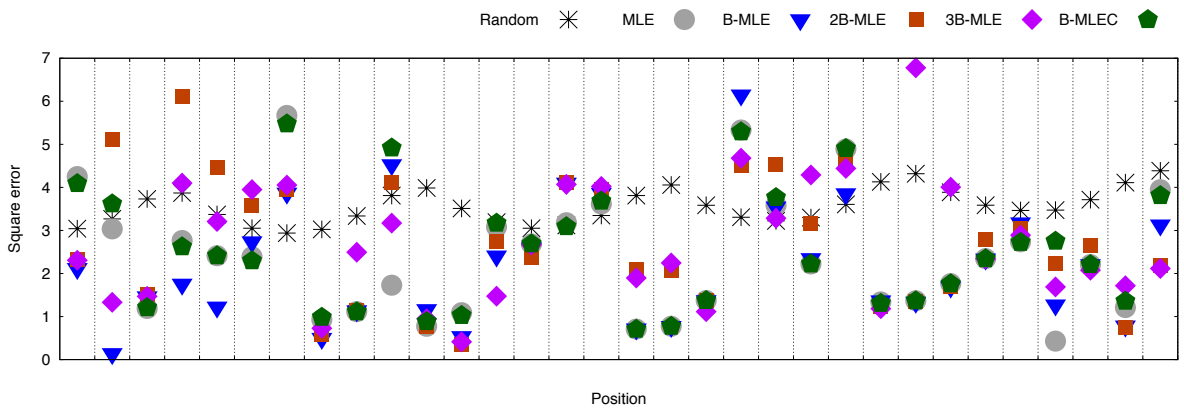


(c) Testbed #3

Figure 4.24: Relative accuracies of the random localization, MLE and B-MLE on the three testbeds and for two learning sets

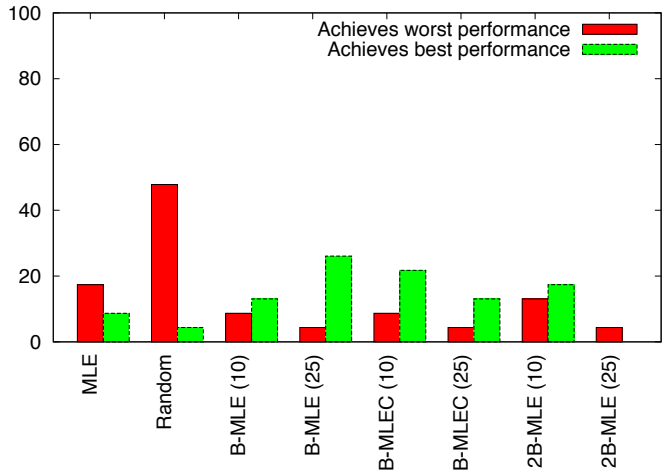


(a) Smaller learning set

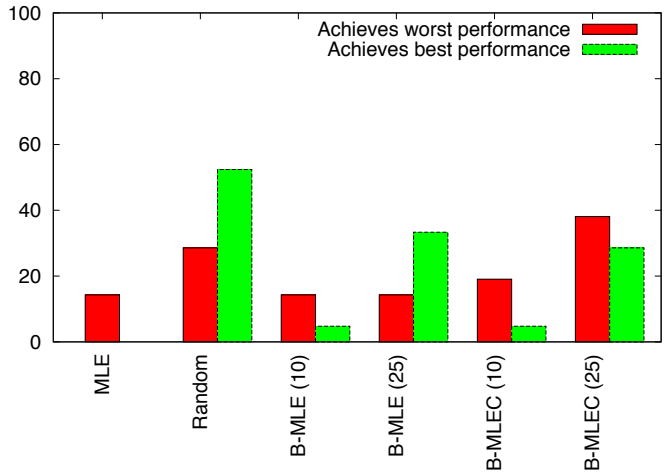


(b) Larger learning set

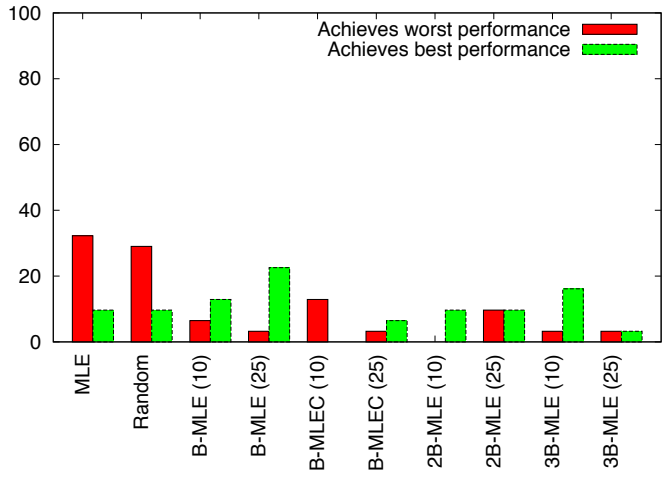
Figure 4.25: Relative accuracies of all strategies on testbed #3 for both learning sets



(a) Testbed 1



(b) Testbed 2



(c) Testbed 3

Figure 4.26: Percentages of positions for which each strategy achieves the best/worst performance of all

4.4 Concluding remarks on biased methods

We presented an algorithm and a few variations to reduce the effects of aberrant landmarks before performing a maximum likelihood estimator (MLE) in order to evaluate a mobile node position in a wireless network. Excluding data coming from one or several landmarks intends on alleviating effects due to multipath propagation. We compared over three different testbeds, algorithms that replace data coming from some landmarks with a constant bias. The landmarks selection is first performed by maximizing the likelihood in the resulting subset of landmarks. We compare, in this situation, the cases when one, two or three landmarks are abnormals with the unbiased MLE. We then evaluate a heuristic approach that does not need to examine all possible subsets of landmarks but selects abnormals landmarks based on a threshold on the bias.

Results do not privilege one strategy over the others. Depending on the testbed configuration, on the variability in the environment and on the time passed estimating propagation parameters (reflected through the *learning data set*), one method can lead to a better accuracy than the others. These experiments showed however that, in the general case, this technique can improve accuracy when compared to classical MLE, ultimately leading to leave only the necessary number of landmarks active. For larger testbeds, the exclusion of a single landmark proves to be the most efficient strategy, while in smaller testbeds, the heuristic algorithm improves results. Considering the global results, it seems that smaller testbeds are more difficult to address than larger testbeds, which means that the effect of the walls, even when they are non-concrete walls is noticeable.

The location is decided instantaneously and thus support the node mobility. Localization is performed on time that the mobile node moves from one location (e.g. #1) to another location (e.g. #2) in 1 second for trilateration and between 1 to 3 seconds for B-MLE, as it can take time sometimes to B-MLE to investigate the position where the target node is located by searching the global maximum of the targeted environment. The average number of RSSI that can be collected a minute is about 140 with a total number of 8 landmarks. With trilateration, there may be some case in which the mobile node cannot obtain the RSSI from some landmarks, because nodes are not close to the mobile and thus their data cannot be received to reach the practical minimum number of the landmarks needed to localize.

This means, our algorithm is robust in dynamic environment where signal condition change, due to shadowing, multipath fading, as what is important is to select the abnormal landmarks adequately rather than a selection of RSSI values from a landmark node, which is used in most localization system in the literature by defining an RSSI threshold in beyond which RSSI values are rejected from a mobile node.

In the following, we will evaluate the use of a *two-modes* of Gaussian Mixture Model (GMM) to reduce the effect of outlier measurements. GMM [27] is a statistical model that models an unknown random variable as the sum of Gaussian random variables. With this method, it is possible to determine the composition function over a training set and to utilize the resulting expression to detect and filter aberrant patterns. We evaluate the accuracy of this method and compare it with regular trilateration and with our previous proposal: B-MLE [186] over a few sets of experimental measurements.

We first introduce the *two-modes* of Gaussian Mixture Models approach in Section 4.5 before presenting an excerpt of the related literature in the same Section. We then describe in Section 4.5.1 the concept of a *two-modes* Gaussian Mixture Model based indoor localization algorithm. Section 4.5.2 provides details about the testbeds we acquired measurements from. The same section presents our experimental results. Conclusions are summarized in Section 4.5.3.

4.5 Why a *two-modes* of Gaussian Mixture Models?

Mixture models [187] are a well known technique in statistics. It can be understood as a weighted sum of laws that provide approximations for a multi-modal distribution. It is used when one analytic distribution does not accurately model the situation. Honary *et al.* [162] use an offline clustering process based on GMM to identify in which room of a building a GSM phone is located. This positioning algorithm is compared to k-Nearest Neighbor (k-NN) based positioning. In an offline phase, a maximum likelihood estimator is used and the GMM algorithm is applied recursively until all GMM parameters converge for each class of the calibration data. Then, an online classification process uses the optimal parameters calculated during the offline process to identify the probability to which the real time data samples belong to the GMM clusters. The authors experimental results have shown that both classification techniques the k-NN approach and the proposed GMM algorithm can provide equally accurate results using tags transmitting at 433 MHz with 3 infrared beacons that are placed in three different rooms. Goswami *et al.* [161] propose to use packet sniffing devices in a wireless LAN to learn parameters of a GMM algorithm. This algorithm, called WiGEM can localize on a building map different WiFi devices that work at different transmit power levels and it performs better than RF map-based techniques.

Such works, even though they require the help of an external element to perform computation, show that GMM is a good candidate to enhance localization accuracy in wireless networks. In the rest of this paper, we consider GMM in a similar way to the approach presented in [162] without an offline clustering process. We do not deal neither with mapping nor using an extended Kalman filter as in [188]. We only use *two-modes* in mixture model: one for the direct path and another one for the outliers. The likelihood of a sample x for Gaussian Mixture Models is given by:

$$F(x) = \sum_{l=1}^{K} \pi_l f_l(x) \tag{4.43}$$

with $\sum_{l=1}^{K} \pi_l = 1$ and $f_l()$ can be any density and the $f_l()$ should not be from the same family. Gaussian Mixture Models (GMM) is statistical Machine Learning approach that have been already used successfully. GMM assumes linear super-position of Gaussians distributions. It can be applied to data where observations emanate from various groups with unknown group affiliations. In a finite mixture model the distribution of the random variable x has a known form (e.g. Normal). However, the values of the parameters of the distribution are unknown. GMM are universal approximates of densities and given a set of data points $x = x_1, \ldots, x_N$, the mixing probability given a mixture of K Gaussians and π_l, the weight of Gaussian l.

$$P(x) = \sum_{l=1}^{K} \pi_l f_l(x \backslash \mu_l, \epsilon_l) \tag{4.44}$$

$$f_l(x \backslash \mu_l, \epsilon_l) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|\epsilon_l|}} e^{(-\frac{1}{2}(x-\mu_l)^t \epsilon_l^{-1}(x-\mu_l))} \tag{4.45}$$

Where d is the dimension of x, μ is the mean and ϵ is the covariance matrix. Moreover, when considering multivariate Gaussian in one dimension ($d = 1$) with mean μ_l and variance σ_l^2.

$$f_l(x \backslash \mu_l, \epsilon_l) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{l}{2\sigma_l^2}(x-\mu_l)^2}$$

Finite mixture models with a fixed number of components are usually estimated with the expectation-maximization (EM) algorithm within a maximum likelihood framework [189]. The idea behind EM algorithm is to introduce a hidden variable in order to simplify the maximization of Equation (4.44). The distribution and the value of the parameters of that hidden variable are then estimated. Finally, the joint distribution of the data and the hidden variable are maximized. In other words, the EM algorithm computes in iterative way the expected maximum likelihood estimate of the whole distribution and then maximizes this computed likelihood of complete distribution.

If θ represents the parameters of the distribution, the maximum is given by:

$$\theta = \arg\max_\theta p(x_1, \ldots, x_N \backslash \theta) = \arg\max_\theta \prod_{i=1}^{N} p(x_i \backslash \theta)$$

Works from Lauer *et al.* [27] extended the use of GMM to describe the pattern's distribution. This so-called *two-modes* GMM use a two-level approach which analyzes the training patterns using widespread outlier distribution to model the existence of outliers. This widespread outlier distribution models the pattern distribution as the composition of a big percentage of normal patterns and a small proportion of corrupted patterns without any need to know the exact number of outliers, neither which training patterns are outlying, nor the exact number of anomalous patterns. The estimation of the parameters of the complete mixture can be done by the EM algorithm. The Em-algorithm locally maximizes the likelihood-function and that the existence of the outlier component bounds the log-likelihood below.

As mentioned, in order to improve indoor positioning and to find the distribution that acquired data (samples) come from among outliers, this GMM approach has been adopted using two modes. Given the data set acquired by a blind mobile node, the idea is to estimate its most probable location by finding the corresponding parameters of the direct path component which is separated to the outliers one.

4.5.1 GMM-MLE localization algorithm approach

In the previous approaches we consider that the path loss, expressed in dB, follows a linear model w.r.t. the log-distance and we have assumed that the model noise is Gaussian. This model is not (well) appropriate in the presence of outliers. In such cases a well-known approach consists of using a Gaussian Mixture Model (GMM) of two modes [27], one for the direct path and the other one to trap the outliers. GMM estimates the distribution of the Normal patterns from training data which contain outliers. We recall the expression of the GMM assuming two modes and N independent, identically-distributed (i.i.d.) observations $X_{n=1:N}$. The log-likelihood function writes:

$$\log p(\mathrm{RSSI}_{1:N}; \theta) = \sum_{n=1}^{N} \log \sum_{k=1}^{2} \pi_k f_k(\mathrm{RSSI}_n; \beta_{0,k}, \beta_{1,k}, \sigma_k) \tag{4.46}$$

where $p(\mathrm{RSSI}_{1:N}; \theta)$ denotes the probability density from the parameters θ, $\theta = (\pi_1, \pi_2; \beta_{0,2}, \beta_{1,2}, \sigma_1; \sigma_2)$ with $\pi_1 + \pi_2 = 1$, and

$$f_1(r; \sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2} r^2}$$

$$f_2(r; \beta_{0,2}, \beta_{1,2}, \sigma_2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(r - \beta_{0,2} - \beta_{1,2} \log_{10}(d_2(x,y)))^2}$$

Table 4.8: Localization error in the LINCS testbed

	Data Set	Average	Median	STD
GMM-MLE	A	1.11 m	1.00 m	0.62
Trilateration	A	1.30 m	1.16 m	0.79
B-MLE	A	1.39 m	1.29 m	0.85
GMM-MLE	B	1.12 m	0.86 m	0.75
Trilateration	B	1.62 m	1.58 m	0.87
B-MLE	B	1.34 m	1.20 m	0.92

Table 4.9: Localization error in the Rammus testbed

	Data Set	MSE	Median	STD
GMM-MLE	A	3.79 m	3.67 m	1.95
Trilateration	A	2.06 m	2.01 m	0.98
B-MLE	A	2.73 m	2.27 m	1.93
GMM-MLE	B	2.37 m	1.88 m	1.62
Trilateration	B	2.36 m	2.13 m	0.92
B-MLE	B	2.20 m	1.94 m	1.40

The likelihood estimate of θ have been obtained by the Expectation-Maximization algorithm [189]. As the maximum likelihood estimator is applied on the function p expressed in Equation (4.46); and as the *logarithm* is a monotonic function, we can apply MLE on the function *log* p. Then, the estimated coordinates (x, y) is obtained as:

$$(\hat{x}, \hat{y}) \quad = \quad \arg\min_{x,y} \sum_{i=1}^{J} \log p(\mathrm{RSSI}_{i,1:N}; \theta) \tag{4.47}$$

The N RSSI samples are measured from the N emitted messages when considering J landmarks in total.

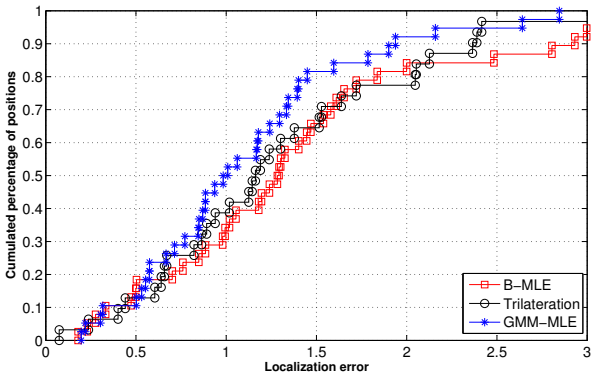
4.5.2 Results and interpretation

We compare the three methods described above (Trilateration, B-MLE and GMM-MLE) on a set of results acquired from two experimental campaigns conducted in the testbeds (#2) and (#3).

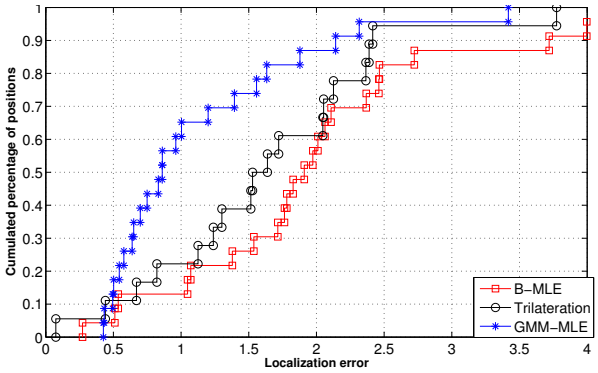
Tables 4.8 and 4.9 report the average, the median and the standard deviation of the positioning error (expressed in meters) for the three algorithms over the testing sets A and B. The testing set represents the number of positions used for algorithms comparison and testing set A is bigger than testing set B. Figure 4.27 shows the CDF of the localization errors for the three algorithms in the four scenarios.

In the smallest testbed (LINCS), both the average values and the graphs show that GMM-MLE achieves a better accuracy. In particular, Figure 4.27(b) shows that about 40 % of the localization errors are below 0.5 m with GMM-MLE. When the learning database is bigger (cf. Figure 4.27(b)), all three algorithms have similar error profiles, even though GMM-MLE has the smallest average and median error.

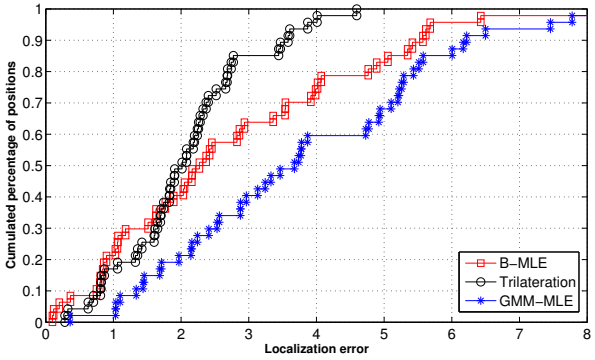
On the largest testbed, though (RAMMUS), trilateration and B-MLE both achieve better results



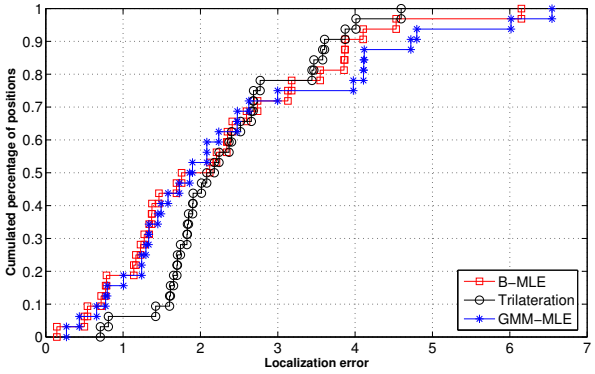
(a) LINCS testbed - testing set A



(b) LINCS testbed - testing set B



(c) Rammus testbed - testing set A



(d) Rammus testbed - testing set B

Figure 4.27: CDF of the distance error for GMM-MLE, B-MLE and trilateration

than GMM-MLE, with a small advantage for trilateration when the learning database is smaller (cf. Figure 4.27(c)). This means that in this scenario, trilateration is good enough and that removing measurements considered as outliers is counter-productive. It may be due to the size of the room, which is slightly larger than the previous testbed. In this case, it would mean that increasing the room size brings experimental conditions closer to a free space situation, and that the threshold between both models should be determined. Nevertheless, the validation of such hypothesis requires additional testings in various rooms.

4.5.3 Concluding remarks on *two-modes* GMM (i.e. GMM-MLE)

In this section, we presented an experiment-based indoor localization method with the use of the imprecise RSSI metric. RSSI is known to be a poor ranging metric in an indoor scenario as it is highly sensitive to signal propagation effects such as multipath. To alleviate the randomness that results from these propagation effects, we evaluated here a method to keep, for the localization process, only the most coherent measurements with the help of a *two-modes* Gaussian Mixture Model. This proposed method does not require any previous complex calibration or area specific training phase. We compare this method with two other candidates: classical trilateration on the whole dataset and a previous proposal based on the introduction of a bias in a maximum likelihood estimator. Results obtained show that the GMM-MLE based method has a good potential, as it improves significantly localization errors in some environments. However, it is not efficient in every scenario and can yield to worse results than trilateration. This means that a localization system should consider this method as one of the potentially relevant tools. However, performing a clever selection between the different available methods requires to identify online the operational scenario and to match it to some criteria such as the room size, or the presence of interfering access points. We believe that such decision could be based on a single node's perception, coming for example from the statistical properties of the dataset.

4.6 Summary and discussion on proposed RSSI-based one-hop indoor localization algorithms

We proposed one-hop localization algorithms to take into account the inconsistencies observed in several indoor environments and highlighted by our measurement campaigns results. Such discrepancies are identified effects introduced by multipath and outliers. We then demonstrated that such proposed algorithms are adaptable to any environment without no need to calibration, no central processing unit and no human intervention. These localization algorithms are indeed able to improve the localization performance in the presence of noise.

With statistics approaches/filtering applied to measured RSSI values and using few landmarks nodes compared to the related work as presented in Table A.2 in appendix A, our proposed one-hop RSSI-based localization algorithms are able to provide sufficient localization accuracy for desired types of application. With our improvements, we are able to follow a person in his home. Our proposed localization algorithms can also enable to locate babies in a nursery, to position accurately a camera in front of a person or to monitor artworks in a museum.

Besides these proposed one-hop localization algorithms, we additionally introduce a distributed localization algorithm to effectively and significantly refine location accuracy of sensor nodes in a wireless sensor network.

The contributions of our work are summarized below. Our proposed algorithms are able to improve the indoor location accuracy performance at about:

- 30% with GMM-MLE compared to B-MLE
- 50% with GMM-MLE over trilateration localization algorithm.
- 40% with B-MLE over MLE
- 20% with B-MLE compared to trilateration
- an average of 70% with DSAL compared to B-MLE
- 85% with DSAL compared to multilateration and MDS

Chapter 5

Proposed RSSI-based distributed localization algorithm

Contents

In previous algorithms, a node can only estimate its own position. Recall that such localization scenario is useful and needed in some application like artworks surveillance in a museum; locate babies in a nursery; localize and track an elderly person in a house, so as to learn their habits and then trigger actions if needed (e.g. for medical reasons). However, in order to be closer to wireless sensor networks-based applications for cooperation purpose, we proposed a distributed localization algorithm. With this proposed distributed stochastic localization algorithm named DSAL, each sensor is able to estimate its own position as well as those of its neighbors. To test its performance, our DSAL localization algorithm is compared with multilateration and MDS-based techniques.

In this chapter, we first describe the principle of data analysis techniques and multilateration-based localization algorithms (*i.e.* geometrical techniques). We also present our distributed stochastic localization algorithm mechanism named (*i.e.* DSAL). Finally we present the DSAL localization accuracy compared to multilateration and MDS-based techniques.

5.1 Data Analysis Techniques : background

5.1.1 Multi Dimensional Scaling (MDS)

Multi Dimensional Scaling (MDS) is a data analysis technique to infer the main dimensions involved within a spatial configuration of objects when the available data is related to some measures of their general similarity or dissimilarity [190]. The data collected for MDS analysis

is arranged as a square matrix named the proximity matrix which describes the overall (dis) similarity of the objects under investigation. The idea behind Multidimensional Scaling for positioning problem is the following: a node computes a relative map (initial local map) according to the relative positions from its immediate neighbors.

This type of method does not require the use of landmarks to find the relative positions of the set of blind nodes unlike the aforementioned methods such as multilateration and Maximum likelihood estimator. Indeed, with MDS, the landmarks are only used if one needs to compute the absolute positions. In this case, It is required to use landmarks. But if an application only requires relative positions, those can be derived without the need of landmarks. Unlike to multilateration, with MDS the coordinates reference is the barycenter of the network formed by all sensors with unknown positions. In [165] different localization techniques are analyzed in depth as long as several examples are shown for different applications.

In [165] the authors proposed two different methods to solve the problem of localization of a set of N unknown positions when data is available under the form of square and symmetric matrix named D_{square} whose elements $\{D_{square}(i,j)\}_{\forall i,j}$ correspond to the set of the $\frac{N(N-1)}{2}$ pairwise square distances between the N nodes.

We denote by $\boldsymbol{Z}$ the matrix containing the positions of each regular sensor node and whose dimensions are $N \times 2$ (two dimensional case). The position of each node i can be defined as $\boldsymbol{Z}(i) = (x_i, y_i)$. As the aim is to find the relative positions $\boldsymbol{Z}$, one found different ways to express the available matrix D_{square} as a function of matrix $\boldsymbol{Z}$. One approach relies matrix D_{square} on a double centered version of Z to obtain the Principal Component subspace (See Chapter 12 in [165]). A second approach defines a scalar function called Stress depending on $\boldsymbol{Z}$ appropriate to minimize iteratively (see Chapter 8 in [165]).

As measured distances are not perfectly known, we assume the following model : a set of N nodes composed of one blind node and virtual landmarks forming a connected graph whose distances are estimated from the RSSI data packets received at each blind node and using the unbiased estimator defined in equation 4.13 for each entrance of the matrix $\hat{d}_{square}$. In such case, we obtained an estimated matrix $\hat{D}_{square}$ and both methods seek to minimize the corresponding mean square error between the estimated and the true matrix. The two proposed approaches in the literature are detailed in [165] and can be summarized as follows :

1. Classic: Principal Component Analysis (PCA)

   This first method seeks to find the principal subspace related to the matrix $\hat{D}_{square}$ (theory of Principal Component Analysis) through its corresponding eigen decomposition. As matrix D_{square} can be written under the form of the unknown positions $\boldsymbol{Z}\boldsymbol{Z}^t$, the method is defined as the following minimization problem such :

$$\min_{\boldsymbol{X}} ||\boldsymbol{X}\boldsymbol{X}^t + \frac{1}{2} J_\perp \hat{D}_{square} J_\perp||^2 \tag{5.1}$$
$$\text{where: } J_\perp = I - 1/N\mathbf{1}\mathbf{1}^t$$

   Equation 5.1 aims to minimize the mean square error between the true unknown distances D_{square} which can be written as a product of positions $\boldsymbol{X}\boldsymbol{X}^t$, and the estimated squared distances obtained with the matrix $\hat{D}_{square}$.

   Note that matrix $\boldsymbol{J}_\perp$ defined as $J_\perp = \boldsymbol{I} - 1/N\mathbf{1}\mathbf{1}^t$ denotes the centralized operation $J_\perp \boldsymbol{Z}$ which means to apply a translation of the N unknown absolute node positions (having

a network configuration) to its corresponding centroid $1/N \sum_{i=1}^{N} \boldsymbol{Z}(i)$. After that, we define the N unknown relative node positions as the matrix $\boldsymbol{X}$ whose dimensions are $N \times 2$. We recall that $\boldsymbol{I}$ is the identity matrix. It is shown in [165] that a solution to equation 5.1 corresponds to find the eigenvalues decomposition (using PCA) to the double-centered matrix $\frac{-1}{2} J_{\perp} \hat{D}_{square}$. In fact, matrix D_{square} can be written under the form : $c\mathbf{1}^t + \mathbf{1}c^t - 2\boldsymbol{Z}\boldsymbol{Z}^t$, where vector c has dimensions $N \times 1$ and is defined as $c = (c(1), \ldots, c(N))$ whose component i is $c(i) = ||\boldsymbol{Z}(i)||^2 = x_i^2 + y_i^2$. Note that vector $\mathbf{1}$ verifies $J_{\perp}\mathbf{1} = 0$ and $\mathbf{1}^t J_{\perp} = 0$. Then, the double centering operation applied to D_{square} gives one of the possible set of relatives nodes positions $\boldsymbol{X}$ as :

$$\frac{-1}{2} J_{\perp} \, D_{square} \, J_{\perp} = \frac{-1}{2} J_{\perp} (c\mathbf{1}^t + \mathbf{1}c^t - 2\boldsymbol{Z}\boldsymbol{Z}^t) J_{\perp} = J_{\perp}\boldsymbol{Z}\boldsymbol{Z}^t J_{\perp} = \boldsymbol{X}\boldsymbol{X}^t \qquad (5.2)$$

Note also that matrix $\boldsymbol{X}$ is equal to $J_{\perp}\boldsymbol{Z}$ then it contains the relative positions to the centroid of the network configuration, *i.e.*, the unknown absolute node positions $\boldsymbol{Z}$ centered at the network centroid defined as $\boldsymbol{Z}_c = (1/N\mathbf{1}^t\boldsymbol{Z}_1, 1/N\mathbf{1}^t\boldsymbol{Z}_2)$, then $\boldsymbol{X} = \boldsymbol{Z} - \mathbf{1}\boldsymbol{Z}_c$.

As the resulting matrix $\boldsymbol{X}\boldsymbol{X}^t$ with dimensions $N \times N$, has rank 2, it can be found the first two associated eigenvalues (λ_1 and λ_2) and eigenvectors ($\boldsymbol{V}_1$ and $\boldsymbol{V}_2$) written as:

$$\boldsymbol{X}\boldsymbol{X}^t = V\Lambda V^t = V\sqrt{\Lambda}\sqrt{\Lambda}V^t \qquad (5.3)$$

$$\text{then} \quad : \quad \boldsymbol{X} = V\sqrt{\Lambda} = [\sqrt{\lambda_1}\boldsymbol{V}_1, \sqrt{\lambda_2}\boldsymbol{V}_2] \qquad (5.4)$$

Regarding the double centering operation in equation 5.2, it could be computed in a centralized way. One sensor node i is chosen to collect all the $\frac{N(N-1)}{2}$ pairwise distances (*i.e.* distances between pairs of N sensor nodes). It simply requires for sensor node i to measure RSSI values from data sent by their $N-1$ immediate neighbors to compute estimated $\{d_{i,1}, \ldots, d_{i,N}\}$; then to communicate with those neighbors at the end to collect the rest of the estimated distances to build the matrix $\hat{D}_{square}$ and finally to compute the low rank matrix $\boldsymbol{X}\boldsymbol{X}^t$. This step involves a communication cost of $\frac{N(N-1)}{2}$ and the number of operations scales $\frac{N(N-1)}{2}$.

The second centralized computation concerns the eigenvalues decomposition in equation 5.3 and it can be implemented with the iterative method of *Power Method* (PM, p.405 in [191]) or the *QR-factorization* (p.114 in [192]). It involves N^2 number of operations in equation 5.3 to compute each eigenvector. Since we first estimate the square of the distance with a variance value defined in equation 4.12, this method will be suitable for position estimation within networks with a lower or medium number of sensors being well connected.

In [167], a localization method based on this centralized approach (MDS-MAP) is used in a wireless sensor network in which sensors are not necessarily all connected with each other and have different connectivity (number of neighbors). A first step using *shortest path computation* is used to estimate the matrix D_{square}. The bound of the algorithm error are calculated and detailed in [168].

Later on, in [193], the authors proposed to use the same method locally, by dividing the network of between 100 and 200 sensor nodes and getting local maps of neighborhoods closest sensor nodes whose positions are estimated by MDS. A final step of joint of local maps is proposed for the global map. For mobile networks or networks with links between

sensor nodes vary with time, the methods that can implement the MDS in a distributed manner is considered [170].

2. Modern MDS ([165], Chapter 8) or Modern Majorization Problem (SMACOF)

   The method known as SMACOF is based on Scaling by Majorizing a Complicated Function. The majorization problem is formulated as a minimization problem of the function $\sigma_r(\boldsymbol{X})$ known under the name of *raw Stress* that expresses the mean square error between the true distances between each pair of nodes i and j. $d_{i,j} = \sqrt{(z_1(i) - z_1(j))^2 + (z_2(i) - z_2(j))^2}$ and the measured distances $\{\hat{d}_{i,j}\}_{\forall i,j}$ as follows:

$$\hat{\boldsymbol{X}} = \min_{\boldsymbol{X}} \sigma_r(\boldsymbol{X}) = \min_{\boldsymbol{X}} \sum_{j>i} w_{ij}(\hat{d}_{i,j} - d_{i,j}(\boldsymbol{X}))^2 \tag{5.5}$$
$$\text{where : } d_{i,j}(\boldsymbol{X}) = \sqrt{(X_1(i) - X_1(j))^2 + (X_2(i) - X_2(j))^2}$$

   As the optimization problem of equation 5.5 concerns the minimization of a non-trivial function due to its locals minima, an auxiliary quadratic function is used to obtain the expected global minimum iteratively. The general method is appeared in [194] where it is described the iterative majorization method. The SMACOF algorithm is later proposed in [166] to adapt to this optimization problem.

   In [165] (Chapter 8) the solution to the equation 5.5 is implemented with an iterative method based on the quadratic majorization using an auxiliary function whose gradient function can be easily computed. It is shown that the solution that cancels the gradient is obtained from the *Guttman Transformation* (TG).

   It is then applied to each estimate $\boldsymbol{X}^k$ at iteration k (cf. the summary of the algorithm on the page 191 of the Chapter 8 in [165]). After an initializing $\boldsymbol{X}^0$ and $\sigma_r(\boldsymbol{X}^0)$ step the algorithm continues as follow:

   i) Increments in an iteration k.

   ii) Calculates the *Guttman Transformation* (TG) of $\boldsymbol{X}^k$.

   iii) Returns to the step `i)` if $\sigma_r(\boldsymbol{X}^{k-1}) - \sigma_r(\boldsymbol{X}^k) > \epsilon$ and $k \leq$ maximum number of iterations.

   The obtained solution results in a configuration network centered to the barycenter (centroid) as in the previous method. Thus, a version of the equation 5.3 is obtained for a given rotation. The algorithm can be implemented in a centralized manner on any wireless sensor node within the network if it is able to collect all the $\frac{N(N-1)}{2}$ pairwise distances between all sensor nodes. A distributed version is proposed by [125] with a sequential approach where at each iteration the global function to minimize is calculated from the aggregation of the local position estimate of each sensor node which is sent in the network through an update message.

5.1.2 SemiDefinite Programming (SDP)

Semidefinite programming (SDP) localization mechanism is a centralized localization solution based on convex optimization. In general, with semidefinite programming the measurements (distance or connectivity) among the wireless nodes in the network are modeled

as a set of convex constraints to solve by the localization optimization problem. More precisely, in this type of method, the solution is found from a convex optimization problem in which the system has a linear structure with some constraints on the unknown variable. This method applied to the localization problem is studied and analyzed by Monatari *et al.* in [168]. They give the bounds of error based on the results of the theory of *Global Rigidity*: for instance, one can consider a sensor network as a connected a graph $G(N, V)$ with N sensor nodes and a set of edges $V = \{(i, j) \,|\, i \sim j\}$ denoting the connections links between sensor nodes. As in the MDS method, a semi definite problem turns to find the Principal Component Analysis (PCA) (decomposition in eigenvectors and eigenvalues) of a square matrix Q under the constraints of being symmetric and defined semi-positive.

Unlike the double centralization in Classical MDS defined by the minimization equation 5.2, the SDP method minimizes the trace of matrix Q ($trace(Q)$) under the appropriate constraints :

However, unlike the double centralization step as in the Principal Component Analysis (PCA) defined by the equation (5.2), the matrix Q of the SDP is calculated as:

$$\hat{Q} = \min_{Q} \text{trace}(Q) = \min_{Q} \ (Q_{11} + \cdots + Q_{NN}) \tag{5.6}$$

$$\text{subject to : } |Q_{ii} + Q_{jj} - 2Q_{ij} - \hat{d}^2_{ij}| \ < \ \Delta \quad \text{if } (i,j) \in V \tag{5.7}$$

$$Q \succeq 0$$

With $trace(Q)$ is the trace of the matrice Q. The *trace* of a matrix is the sum of the elements being in the diagonal of a matrix.

The solution to the equation (5.6) satisfies the structure $\boldsymbol{X}\boldsymbol{X}^t$ as in MDS algorithm and therefore a step as in Principal Component Analysis (i.e. PCA) as defined in the equation (5.3) is used to find the relative positions from the eigenvalues and eigenvectors. The complexity of this method is how to solve a system of up to $\frac{N(N-1)}{2}$ unknowns and equations when N is number of sensors with a limited number of operational resources.

5.2 Implementation of RSSI-based distributed localization algorithms using data analysis techniques

In this section we present simulation and experimental results of the proposed distributed localization algorithm named DSAL and compare such results to those from multilateration, classical MDS and Modern MDS.

5.2.1 Used distance error formula

We recall the expression of the average estimation error (E_m) used to evaluate the performance of localization mechanisms under simulation and when using real sensors data. If we consider a set of N positions to estimate, $\{\boldsymbol{X}(i)\}_{i=1,\ldots,N}$ in which, each $\boldsymbol{X}(i)$ corresponds to the two dimensions (i.e. $2D$) position of the sensor node i and all of these estimated positions $\{\hat{\boldsymbol{X}}(i)\}_{i=1,\ldots,N}$ is defined as E_m:

$$E_m = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\boldsymbol{X}}(i) - \boldsymbol{X}(i)\| \tag{5.8}$$

In the equation (5.8) we have units of distance in meter m for our experiments. To obtain a qualitative and absolute value in percentage, the expression in equation 5.8 can be normalized by a factor that takes into account the dimensions of the area in which evolves the wireless sensor network. For example, consider experiments with a network of sensors on a rectangle of dimensions $Pm \times Qm$. The value of the distance d is normalized by the diagonal of the rectangle to obtain an estimate of the normalized mean error (NME) as:

$$\text{NME} = \frac{1}{\sqrt{P^2+Q^2}}\frac{1}{N}\sum_{i=1}^{N} \|\hat{\boldsymbol{X}}(i) - \boldsymbol{X}(i)\| \tag{5.9}$$

If the Normalized Mean Error named NME is calculated for each position i, it can be defined as $\frac{1}{\sqrt{P^2+Q^2}}\|\hat{\boldsymbol{X}}(i) - \boldsymbol{X}(i)\|$.

5.2.2 Reference: Monte Carlo simulation for generating RSSI data

Monte Carlo algorithm is referred to a statistical technique in which a large quantity of independent-random samples are generated and evaluated under an optimally chosen probabilistic distribution function like Uniform, Normal, Log normal, and so on [195]. The Monte Carlo approach performs the distribution from the generated data sample thus it can determine numerical results which are sampled at random from the input probability distributions. Then, one can derive whether the empirical result matches or not one of the performed distributions within a given confidence level.

In Monte Carlo simulation analysis, each set of random samples is named a realization and it may be generated independently more than hundred or even thousand times. Such realizations are computed each time using a different chosen density probability function from a known distribution. When analyzing the performance of the different localization methods from simulated RSSI data, we draw data samples from several Monte Carlo runs whose distribution are chosen as log-Normal. Simulated RSSI data is used with the implemented algorithms in order to quantify the theoretical accuracy of such localization algorithms. Then, they are compared to the empirical accuracy obtained when using real measured RSSI values from our testbeds. Monte Carlo analysis is performed with MATLAB and Statisticals Toolbox.

5.2.3 Multilateration-based localization algorithms

Localization with multilateration algorithm is first validated by generating RSSI measurements using Matlab (Monte Carlo realization) according to the log-normal shadowing model (cf. equation 4.1). We obtain non-biased estimated square distances as described above in the section 4.2.2.2 with a value of $\eta = 2.1$. Figure 5.1 shows the mean error estimation of N positions in percentage when taking the diagonal of the rectangle $5 \times 8m$ as a normalization term of the error for different noise values levels (σ^2) of the model.

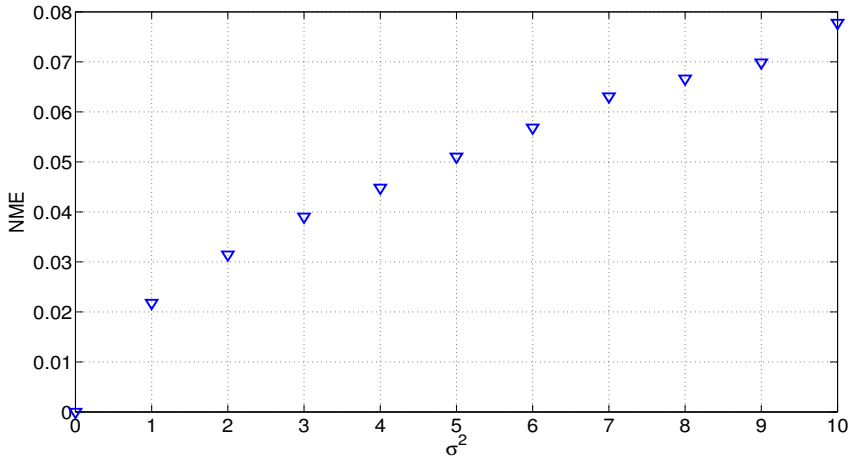


Figure 5.1: Normalized Mean Error (NME) of $N = 12$ estimated positions with multilateration and when using different noise values of the variance of the model: σ^2.

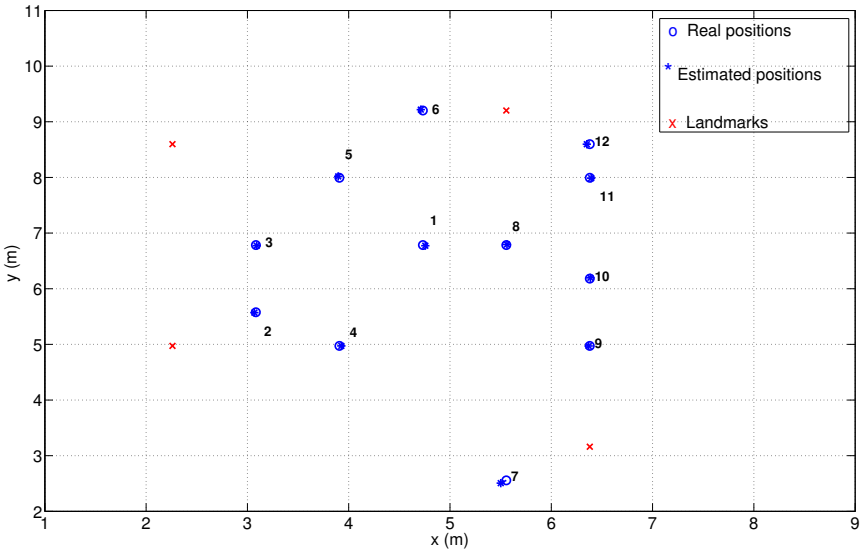
In Figures 5.3(a) and 5.3(b) we show the results achieved with multilateration using one Monte Carlo realization for 12 estimated nodes positions including blind node and virtual landmarks. For such Monte Carlo realization we generated the random square distances using the unbiased estimator expression presented in equation (4.13) with $T = 1$, $\eta = 2.1$ and log-normal random noise with parameter σ^2 equal to 2 and 10. The effect of the variance of the multilateration estimator is analyzed according to the expression in equation (4.27). From this expression the variance depends and increases with distances and σ^2.

The performance of position estimation with multilateration under the log-normal shadowing model assumption is validated using empirical data to estimate the x-coordinate of the blind node identifier 214 position using multilateration as illustrated in Figure 5.4. Empirical data are illustrated according to 1000 Monte Carlo independent realizations (1000 rounds of RSSI data generated with Matlab for 12 unknown positions) with additive noise σ^2 equal to 10. Empirical results are then compared to a Gaussian distribution centered at the real value (real abscissa) of this coordinates with a variance value obtained from the theoretical covariance matrix of the estimator according to the expression of equation (4.27). The Multilateration algorithm runtime for the estimation of 12 unknown nodes positions is 0.47 seconds.

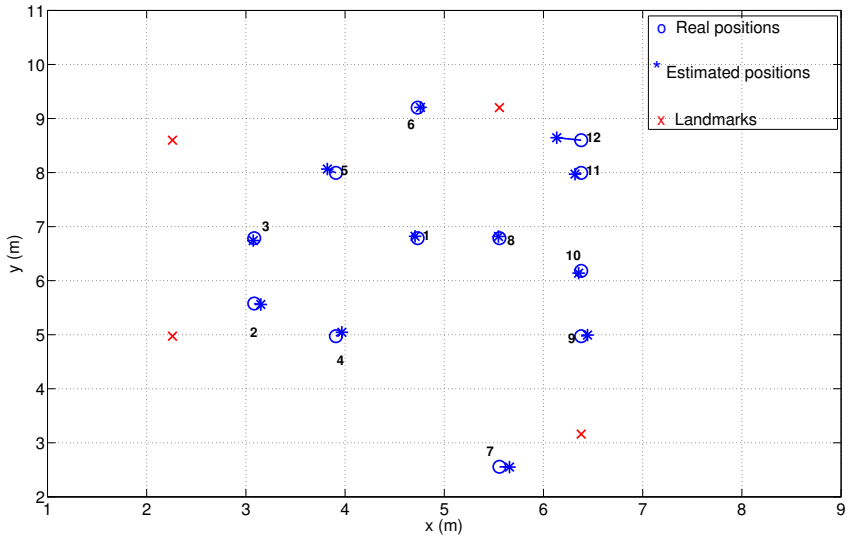
To confirm the theoretical results observed on position estimation with multilateration, we realize 10 independent experiments by collecting RSSI measurements on 12 nodes on a total of 16 wireless sensor nodes from the SensLAB Rennes platform. These 12 are blind nodes and virtual landmarks with unknown positions and 4 remaining nodes are landmarks. Recall that each blind node is considered as a virtual landmark from another blind node point of view.

In Figure 5.5, we illustrate the average estimate of 12 positions for 12 blind nodes with estimated distances from 10 independent experiments. 600 RSSI measures are recorded at each experiment for each blind node position. A blind node can communicate with its neighbors landmarks and virtual landmarks.

From the aforementioned RSSI measures, an average value of $\sigma^2 = 16.37$ is estimated at involved sensor node network selected in the SensLAB Rennes testbed. In this case, the normalized mean error estimation for each of the N sensor nodes are:

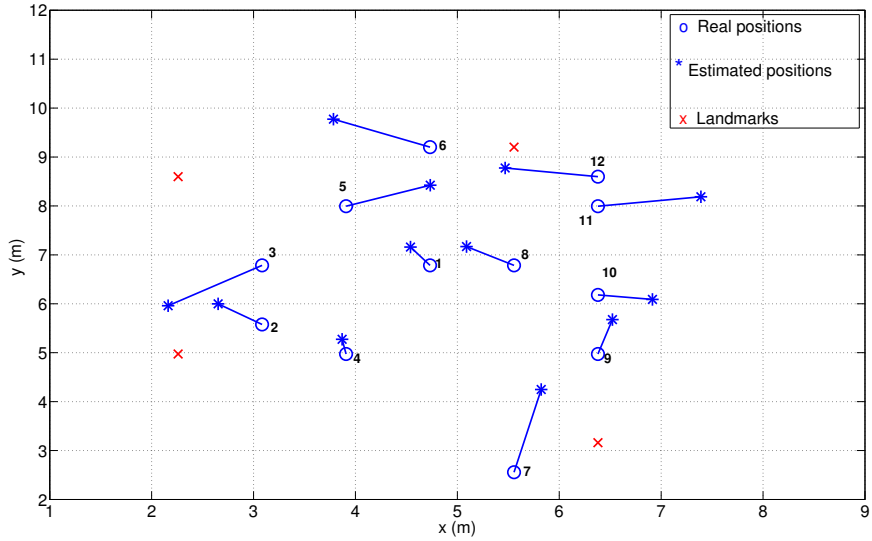


(a) Noise variance : $\sigma^2 = 2$.

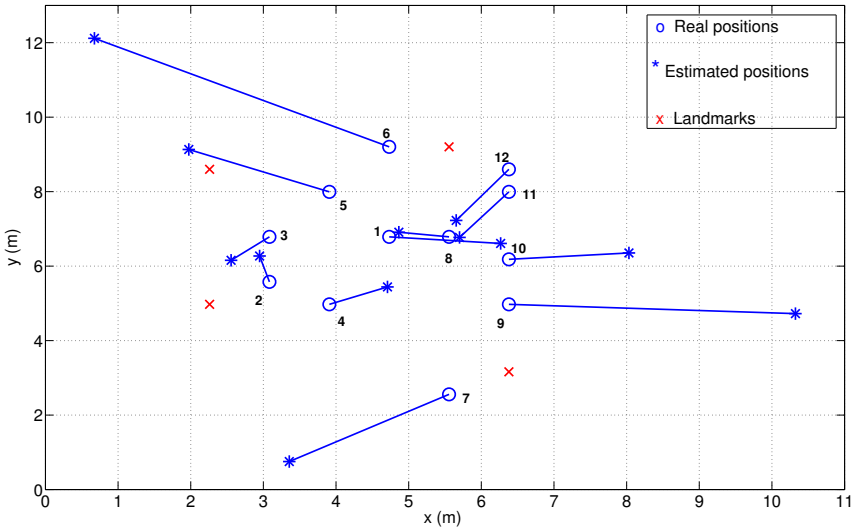


(b) Noise variance : $\sigma^2 = 10$.

Figure 5.2: Average estimated position with multilateration from 1000 independent Monte Carlo realizations for $N = 12$ nodes positions including blind node and virtual landmarks



(a) Noise variance: $\sigma^2 = 2$.



(b) Noise variance: $\sigma^2 = 10$.

Figure 5.3: Estimated positions with multilateration when considering one Monte Carlo realization (RSSI data generated with Matlab)

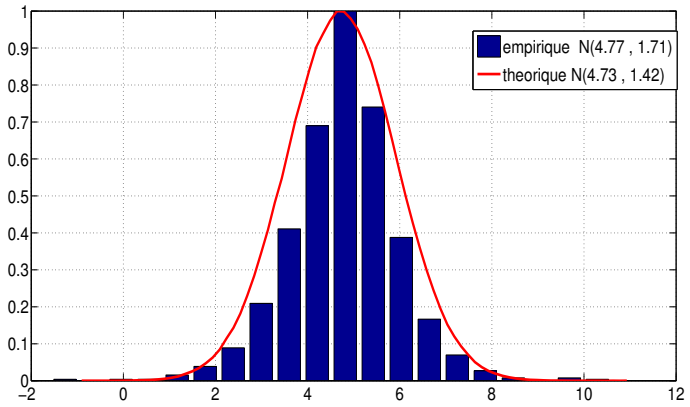


Figure 5.4: Histogram when estimating the abscissa of the blind node (214) when considering empirical and theoretical model with $\sigma^2 = 10$

$$(0.13, 0.2, 0.05, 0.21, 0.2, 0.45, 0.22, 0.21, 0.2, 0.25, 0.25, 0.35)$$

The mean positioning error (normalized mean error along the diagonal of the concerned rectangle) for N sensor nodes under 10 experiments is: $NME = 0.23$.

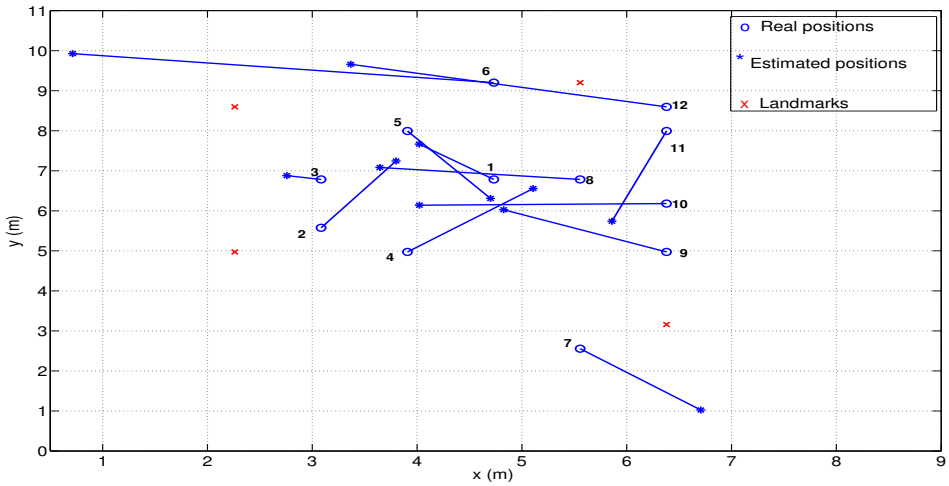


Figure 5.5: Average estimated positions with multilateration for each of the 12 nodes (blind nodes and virtual landmarks). 10 independent experiments using real RSSI measurements are performed at the remote SensLAB Rennes platform. At each experiment, each sensor node collects until 600 packets.

5.2.4 Classic MDS-based localization: Principal Component Analysis (PCA)

We implement Classic MDS-based localization algorithm by analyzing the network configuration of 16 sensor nodes as shown in Figure 4.4. We use the algorithm described in equations (5.2) and (5.3) to calculate the $N = 12$ unknown positions of the blind nodes. We first carry out Monte Carlo simulation by generating in Matlab the values of the squared distances with additive noise and by varying the value of σ^2 between 0 to 10. The purpose of such process is to obtain unbiased estimates of the matrix D_{square}. We then

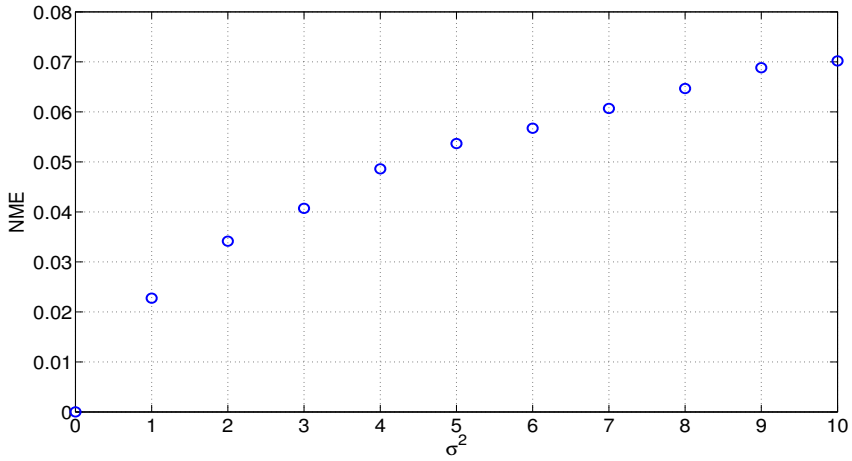


Figure 5.6: Normalized mean error (NME) of Classic MDS-based localization of $N = 12$ estimated positions for different values of σ^2 with Monte Carlo simulation using 1000 Monte Carlo independent realizations (RSSI data generated with Matlab).

illustrat in Figure 5.6 the values of the normalized mean error (NME) (cf. the equation (5.9)) of N positions when considering 1000 Monte Carlo independent realizations (RSSI data generated with Matlab) for each value of σ^2.

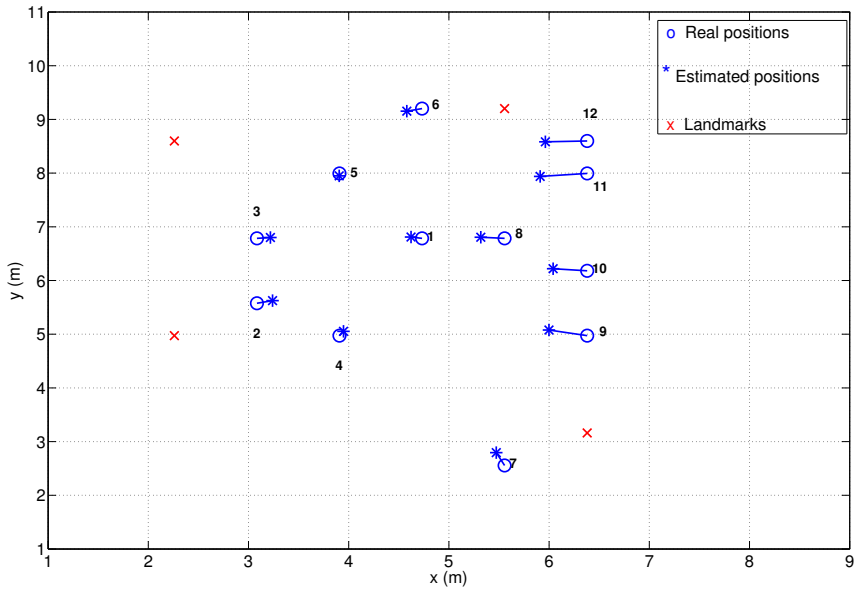
Resulting positions $\hat{\boldsymbol{Z}}$ are estimated with Classic MDS after resetting the relative positions $\boldsymbol{X}$ to the absolute positions $\boldsymbol{Z}$ using $M = 4$ landmarks with the procedure explained in ([165], Chapter 12). The results of the averages positioning errors from 1000 Monte Carlo independent realizations (RSSI data generated with Matlab) are shown in Figure 5.7(a) and 5.7(b). Note that the runtime of the Classic MDS algorithm for the estimation of 12 unknown nodes positions is 1.48 seconds. We observe that the averages estimated positions have a bias proportional to their distance to the centroid of the network and the configuration of the network is maintained.

Figures 5.8(a) and 5.8(b) illustrate the position estimation for a given Monte Carlo realization when σ^2 is equal to 2 and 10.

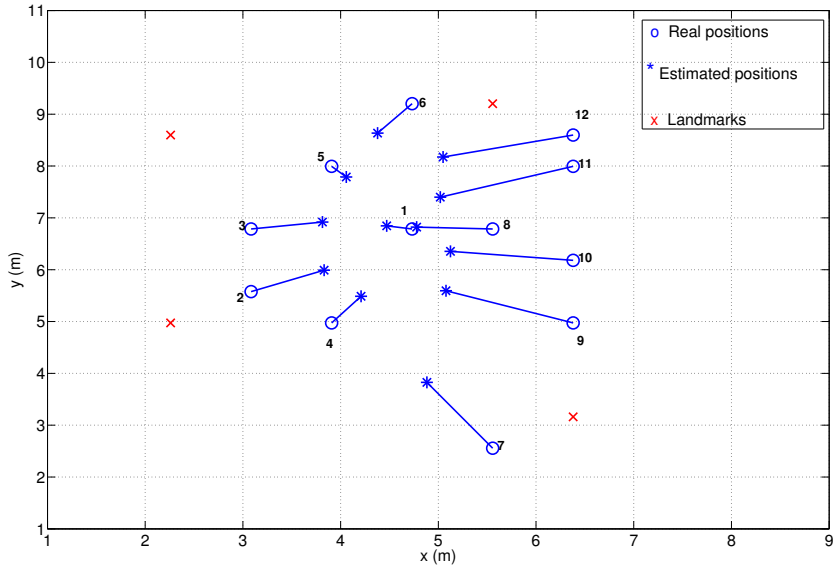
Without noise, the eigenvalues of the matrix H (from which the decomposition of the main components) are equal to $\lambda_1 = 63.25$ and $\lambda_2 = 34.09$. The empirical laws of the first two estimated eigenvalues $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are illustrated in Figure 5.9. As noted in the Section 4.2.1, the maximum likelihood estimator of the distances follows a log-normal distribution depending on the value of σ^2 and thus on the model of the measured RSSI.

The average empirical law of the estimator $\hat{d}_{square}$ is illustrated for the sensor node 1 and for the sensor node 2 as well as for the two values of σ^2. The first sensor node is located close to the centroid. The distances that separate it to the remain sensor nodes within the network is around an average distance. The sensor node 12 is located in an outside corner of the network and the distances to the other sensors are twice the distances reported for the sensor 1. This effect is marked on the variance as we can observe that it depends on the distance to exponent 4. If the average of the square of the distance of the sensor node 1 are computed; $\{d^2_{1,i}\}_{i=1,\ldots,16}$ is equal to 6.15 m^2 (i.e. square meters) and the one for the sensor node 12 is equal to 12.94 m^2.

Experiments are realized on the wireless sensor nodes at SensLAB Rennes platform. We consider 10 independent experiments and the network configuration presented in Figure 4.4. At each experiment, 600 RSSI measures are recorded for each blind sensor node

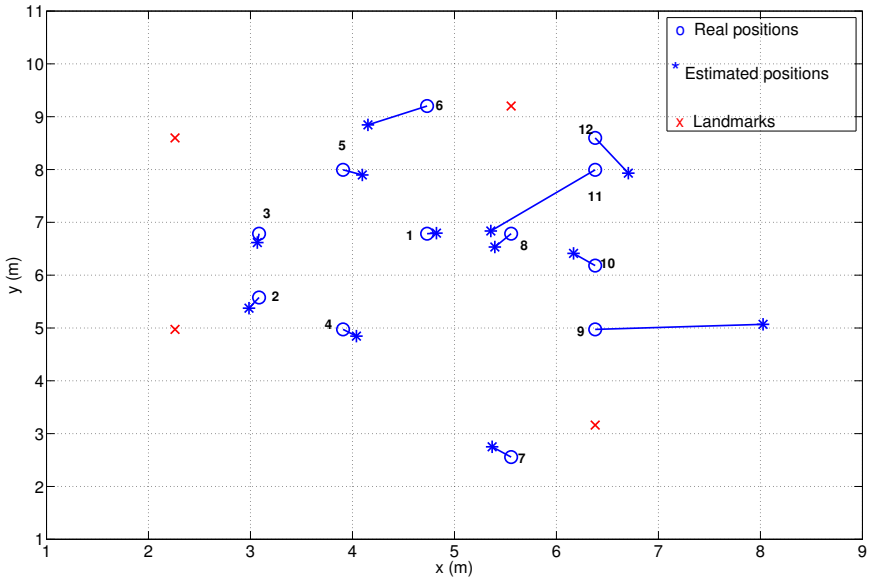


(a) Noise variance: $\sigma^2 = 2$.

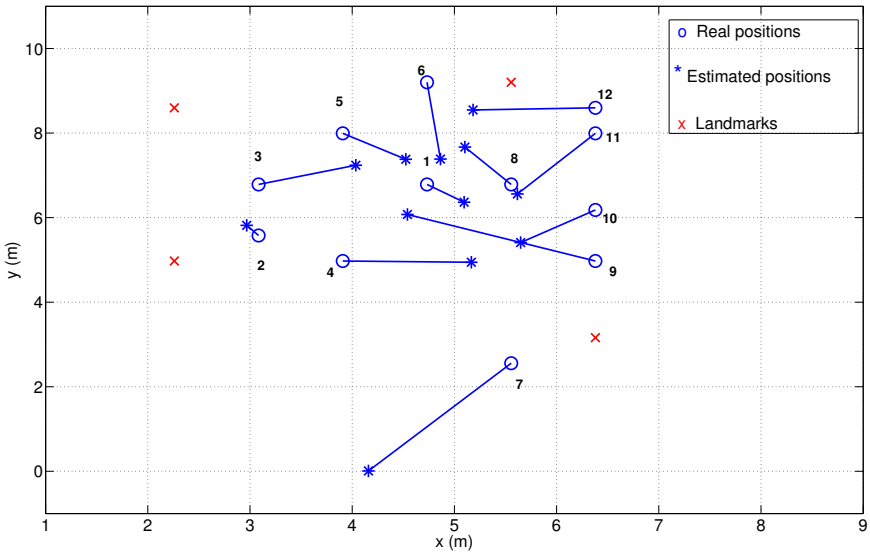


(b) Noise variance: $\sigma^2 = 10$.

Figure 5.7: Positioning error with classical MDS using 1000 independent Monte Carlo realizations (RSSI data generated with Matlab) for $N = 12$ positions



(a) Noise variance: $\sigma^2 = 2$.



(b) Noise variance: $\sigma^2 = 10$.

Figure 5.8: The estimation of $N = 12$ positions with classical MDS-based localization algorithm using one Monte Carlo realization (simulation)

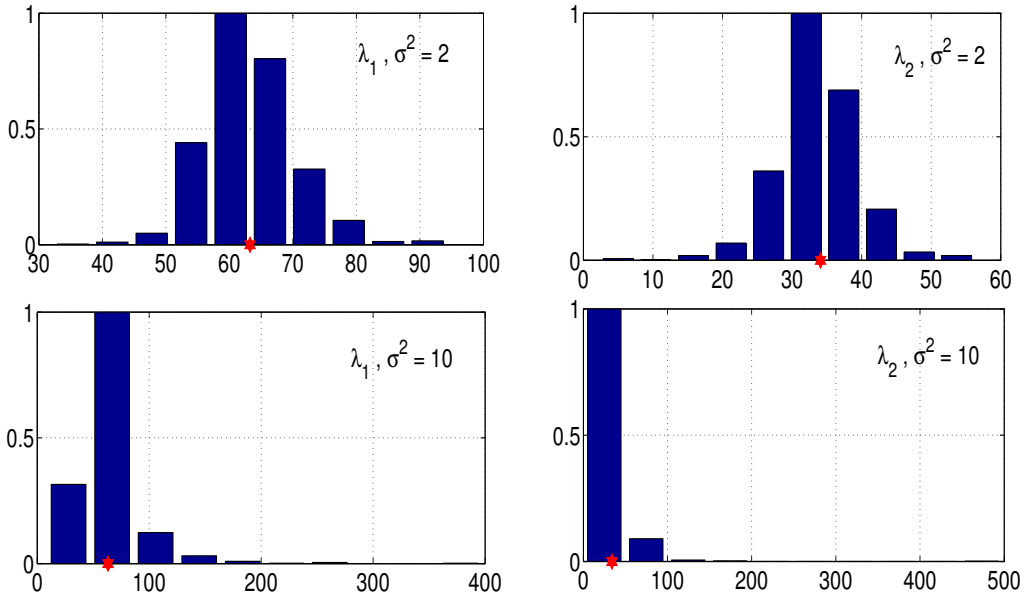


Figure 5.9: Histograms of the first two estimated eigenvalues with classical MDS using Monte Carlo simulation under 1000 independent Monte Carlo realizations (RSSI data generated with Matlab). $\sigma^2 = 2$ (top figures) and $\sigma^2 = 10$ are considered. The true values of λ_1 are marked with (★) (right side) and those of λ_2 are represented at the left side.

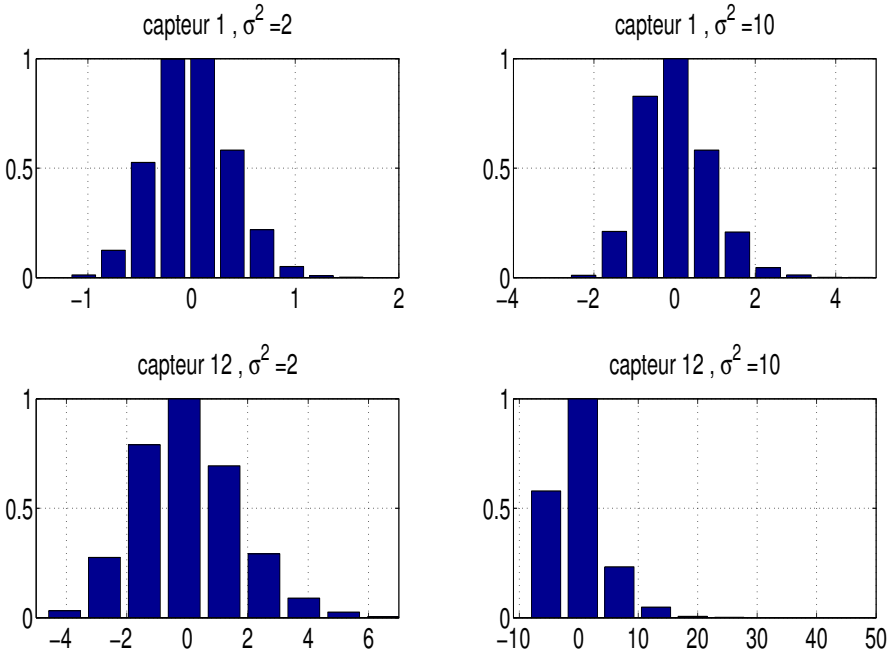


Figure 5.10: Histograms of the average estimated squared distances for the sensor $ID = 1$ (top graphs) and for the sensor node $ID = 12$ (bottom graphs) for values of σ^2 equal to 2 (left) and 10 (right).

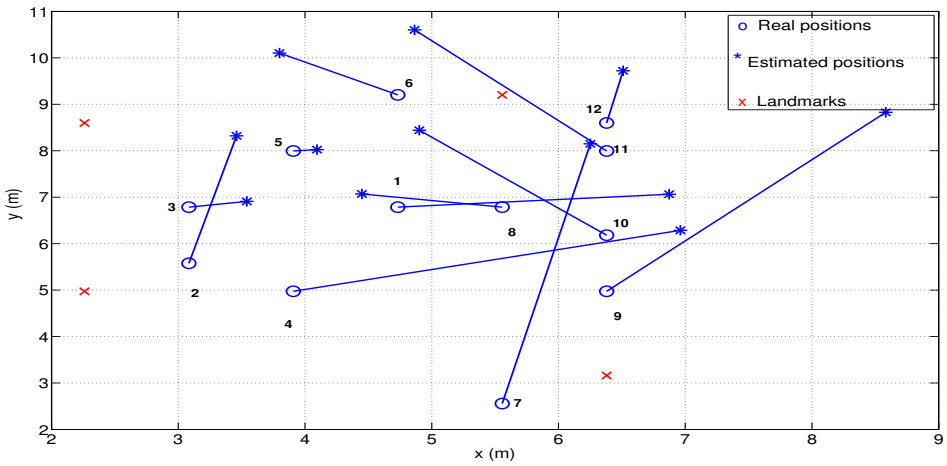


Figure 5.11: Average positions estimate of $N = 12$ positions with classic MDS-based localization from 10 independent experiments realized in 2 hours on sensor nodes at SensLAB Rennes platform.

position. The total duration for all experiments is 2 hours. Note that, in each experiment, due to the layout of the room and disturbances, all the blind sensor nodes were not able to communicate with all landmarks nodes. Figure 5.11 illustrates the average of N estimated positions obtained with the conventional MDS method. A normalized mean error (NME) of 0.3 is obtained. The values of the normalized mean error estimated at each position are:

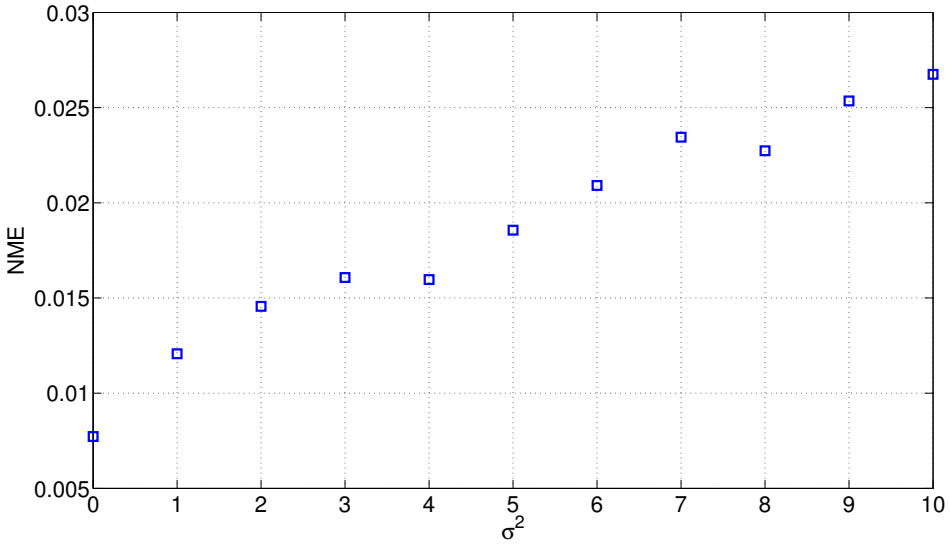
$$(0.23, 0.29, 0.05, 0.35, 0.02, 0.14, 0.59, 0.12, 0.47, 0.29, 0.32, 0.12).$$

The average estimate of the two eigenvalues are equal to $\hat{\lambda}_1 = 42.19$ instead of 63.25 and $\hat{\lambda}_2 = 35.7$ instead of 34.1.

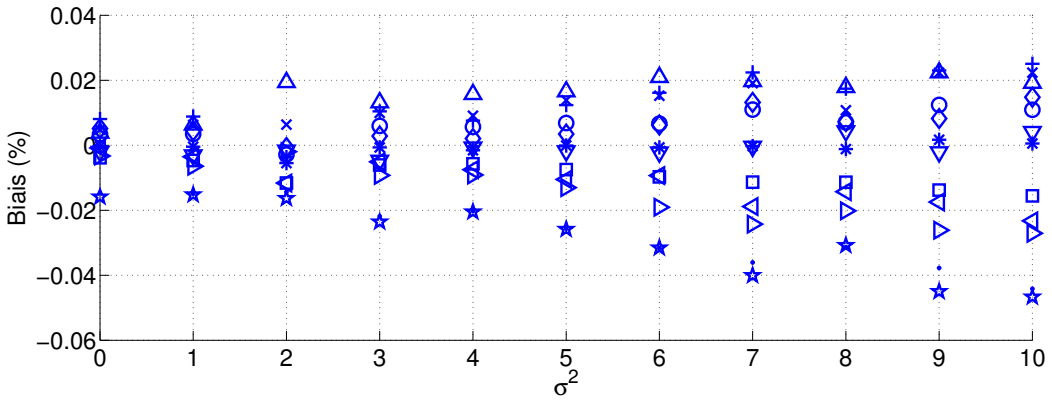
5.2.4.1 Modern MDS-based localization: Modern Majorization Problem (SMACOF)

We analyze as in the previous method the results by considering the network in Figure 4.4 in which we aim to compute the $N = 12$ unknown positions of the blind nodes. First data are simulated in the same manner as above under different noise levels. The initialization point is randomly selected for each sensor node in the network using two uniform components between 0 and 5. In Figure 5.12(a), the normalized average error estimate of N positions is illustrated (cf. the equation (5.9)) using Monte Carlo simulation with 1000 independent realizations (RSSI data generated with Matlab); and for each value of σ^2 between 0 and 10; and with 300 iterations of the SMACOF algorithm. Figure 5.12(b) shows the normalized bias (in the same factor as the expression of the normalized average error estimate in the equation (5.9)) for each estimated position in which the error increases with the value of σ^2.

We observe that the estimated position of the blind sensor node with identifier 1 (node 214): $\boldsymbol{Z}(1) = (4.73, 6.79)$ is the closest to the centroid. Unlike, most external blind nodes locations estimation are slower to converge to the correct solution and the introduced bias is more important. Note that the centroid in this network is $(4.78, 6.52)$). In all achieved results, the estimated location of the bind sensor node with identifier 1 is the one whose the estimation error is the smallest. With the SMACOF algorithm the relative positions



(a) Normalized mean error (NME) of the $N = 12$ estimated positions.



(b) Normalized bias of each blind node.

Figure 5.12: Analyze of positions estimation with the SMACOF algorithm for different values of σ^2 (from 0 to 10)

are obtained at centroid $\boldsymbol{X}$ centered at $(0, 0)$. The algorithm is initialized at points that are close, the minimum is quickly determined and is the same with the true location.

Figures 5.13(a) and 5.13(b) show the average results of the estimated positions when σ^2 is equal to 2 and 10.

Furthermore Figures 5.14(a) and 5.14(b) illustrate the result for one Monte Carlo realization (one round of RSSI data generated with Matlab). However, when considering the RSSI measures from the SensLAB Rennes Platform as shown in Figure 5.15(a), the N positions estimates are obtained using the SMACOF method with random initialization as in the previous case and after 300 iterations. The corresponding runtime for such estimation when $N = 12$ is 1.644 seconds.

For one independent experiment, 6000 RSSI measures are taken at each blind sensor node position for a duration of 2 hours. The normalized mean error (NME) of the N estimated positions reports a value of 0.16 for each position and corresponds to:

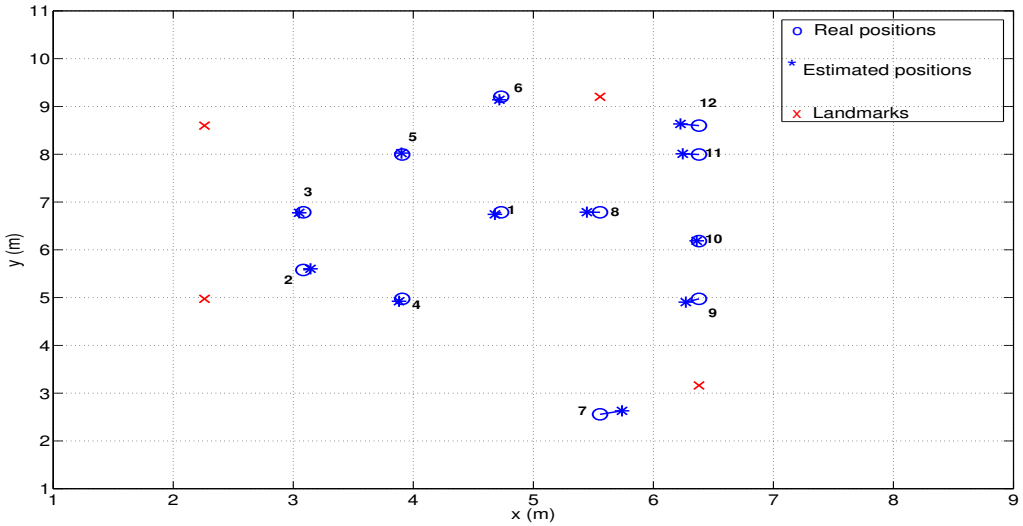
$$(0.2, 0.11, 0.27, 0.09, 0.11, 0.09, 0.16, 0.25, 0.063, 0.41, 0.13, 0.088).$$

The average normalized mean error (NME) is illustrated for 10 independent experiments. At each experiment, each blind sensor node measures 600 RSSI values in a duration of 10 minutes. A value of the normalized mean error (NME) on the 12 estimated positions is 0.25 and for each position the estimated individual errors are:

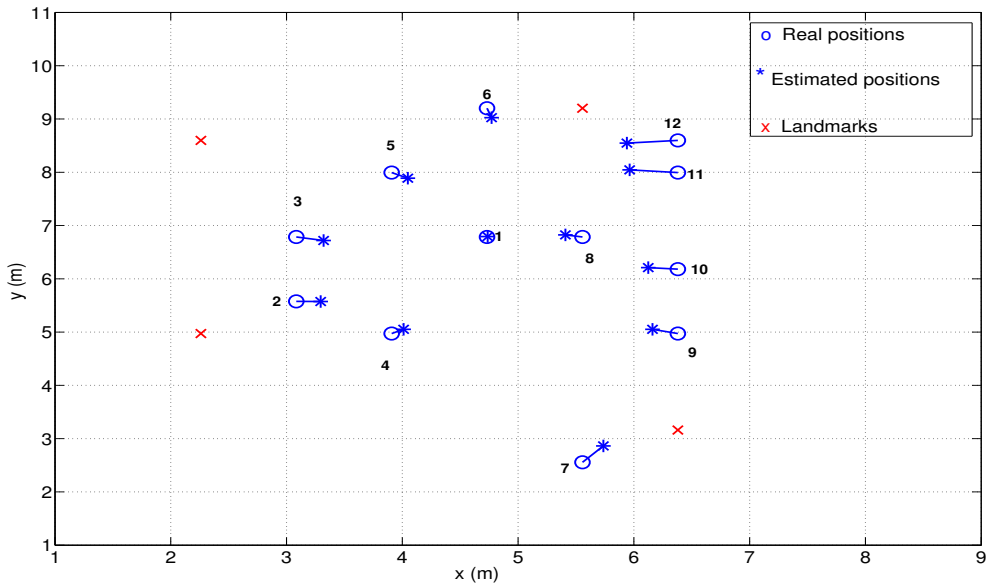
$$(0.19, 0.17, 0.27, 0.23, 0.15, 0.28, 0.51, 0.24, 0.063, 0.41, 0.35, 0.18).$$

5.2.5 On-line Distributed Stochastic Approximation-based localization algorithms (DSAL)

In this section we introduce a new refinement phase based on distributed stochastic approximation to improve the estimated position accuracy in real noisy environments in which the presented methods have a fairly low performance. It has been already discussed an existing general structure to solve the problem of localization of a set of N unknown positions in wireless sensor network (see in [1]). Such structure uses range or distance measures and a number of known positions from a set of M landmarks which are usually placed at the surrounding area of the network and it is usually verified such $M << N$. The common structure of such well-known methods ([29, 30, 89, 193]) is: compute the estimated distances matrix, apply a core algorithm to solve the location problem and finally introduce a refinement phase using a cooperative-based technique to obtain high accurate estimated positions. Such cooperative-based approach uses the temporal estimated positions of the unknown nodes in addition of connectivity information and known landmarks positions [28, 29, 30, 31]. Examples of such refinements phases are : in [89]; the authors computed trilateration step iteratively at each node being declared as a landmark when it becomes well located depending on its confidence level; in [30] they choose a distributed iterative *Kalman* filter to solve the least square minimization problem implemented sequentially throughout the sensors at each iteration to guarantee the global optimum is achieved; and similarly, in [193] the authors solved the least square minimization problem using a centralized gradient descent approach implemented using the *Levenberg-Marquardt* algorithm. Because the non-linear and non-convex nature of the function to minimize, the use of gradient descent approaches is limited to the refinement phase in order to ensure an appropriate initialization first; which depends on the result given by the previous phase, and the global optimum then.

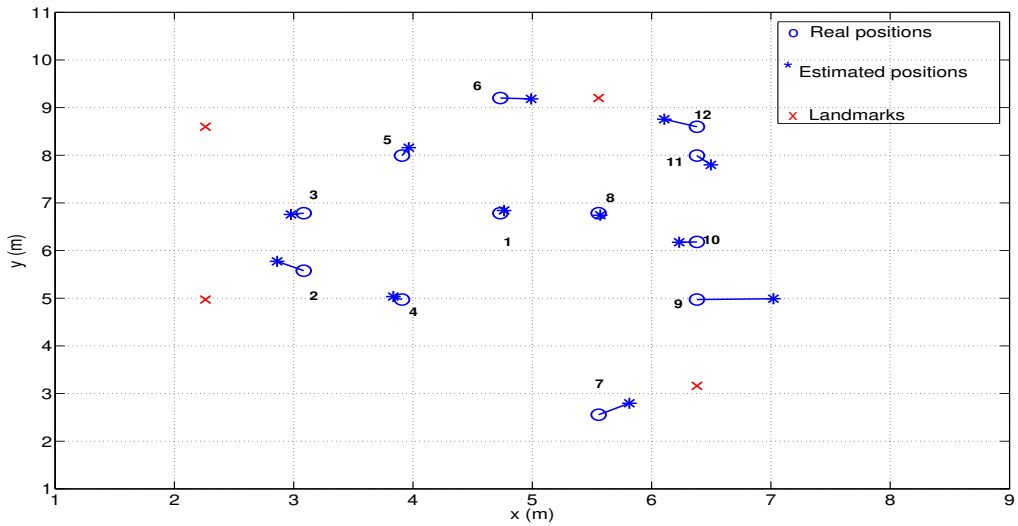


(a) Noise variance: $\sigma^2 = 2$.

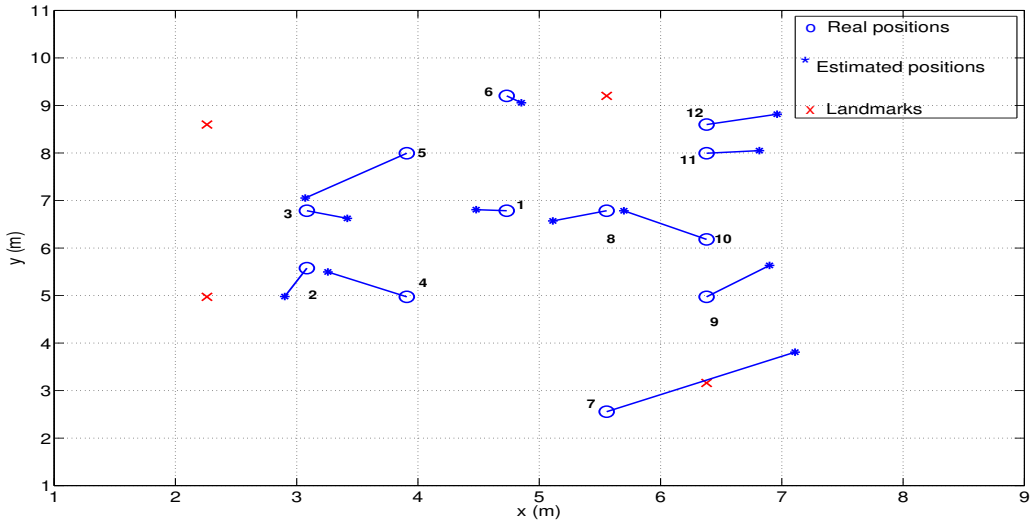


(b) Noise variance: $\sigma^2 = 10$.

Figure 5.13: Average normalized mean error estimation of $N = 12$ unknown positions with the Modern SMACOF algorithm under 1000 independent Monte Carlo realizations (RSSI data generated with Matlab)

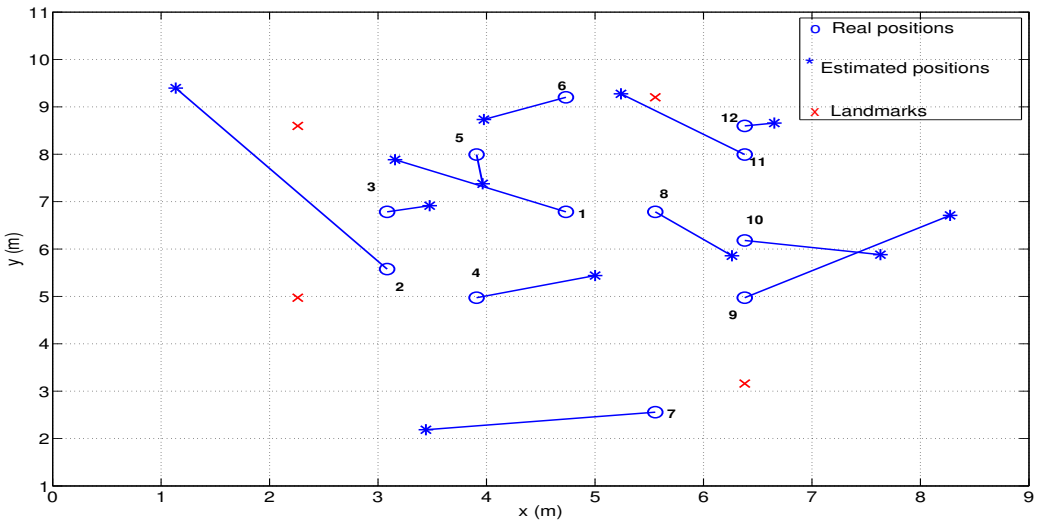


(a) Noise variance: $\sigma^2 = 2$.

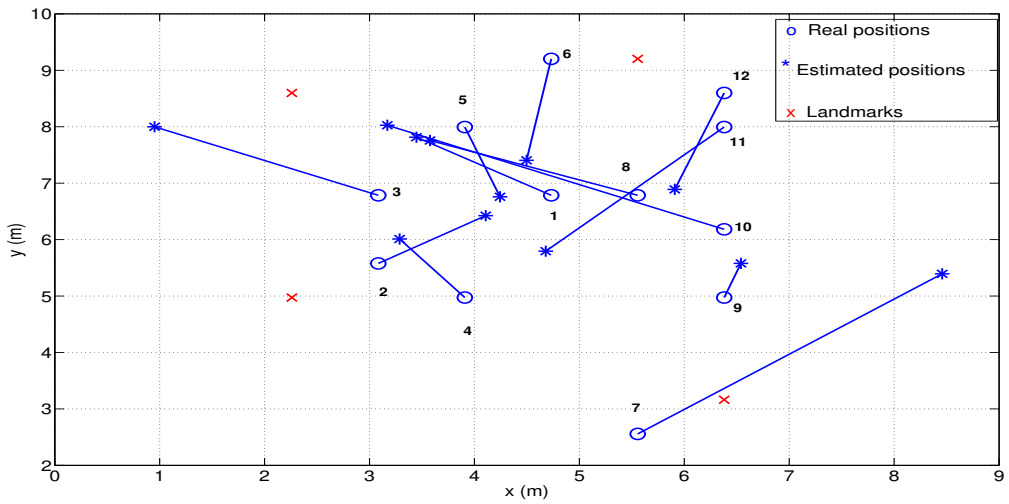


(b) Noise variance: $\sigma^2 = 10$.

Figure 5.14: Estimation of the $N = 12$ positions using SMACOF algorithm



(a) Position estimation of $N = 12$ positions with the SMACOF algorithm for one experiment



(b) Average position estimation for $N = 12$ unknown positions using 10 independent experiments at the Senslab Rennes platform.

Figure 5.15: Results of the SMACOF localization algorithm using RSSI measures from experiments at SensLAB Rennes platform

In our work inspired from the two-phases of above mentioned approaches, we propose to implement an iterative on-line distributed stochastic approximation approach based on [32] which joins two methods : stochastic approximation ([196], [171]) and gossip-based communications ([33]). Such gradient descent approach helps to minimize the mean square error of the estimated distance derived from the noisy measured RSSI from received data. In the literature, we found several works related to the distributed gradient descent where it was firstly proposed in [197], later analyzed in [198]) and proved in both theoretical and practical ways ([199], [200]). Before defining the algorithm, we first detail the notation as follows.

We define the N unknown node positions as the set $\{z_i\}_{i=1,...,N}$ and the set of the M known landmark's positions as $\{\zeta_k\}_{k=1,...,M}$. For any sensor node i, we denote by $\mathcal{N}_i$ the set of sensor nodes $j \neq i$ such that $d_{i,j} = |z_i - z_j| < r$ and $\mathcal{A}_i$ all landmarks k such that $d_{i,k} = |z_i - \zeta_k| < r$. We assume that each sensor node i is at least connected to one landmark, *i.e.* $|\mathcal{A}_i| \geq 1 \ \forall i$. Moreover, we recall that for each sensor node i its position in two dimensional spaces is written as the following two-coordinates vector : $z_i = (x_i, y_i)$. We assume also that any sensor node i can compute the following set of positions which include its own position and the set of its neighboring positions as:

$$\boldsymbol{z}^{(i)} = (z_j)_{j\in\mathcal{N}_i\cup\{i\}} \quad \text{whose two coordinates are:} \quad \boldsymbol{x}^{(i)} = (x_j)_{j\in\mathcal{N}_i\cup\{i\}}\,, \quad \boldsymbol{y}^{(i)} = (y_j)_{j\in\mathcal{N}_i\cup\{i\}}$$

Because the set of pairwise distances $\{d_{i,j}\}_{\forall j\in\mathcal{N}_i}$ and $\{d_{i,k}\}_{\forall k\in\mathcal{A}_i}$ are not perfectly known at each sensor node but they can be estimated from noisy measures of RSSI signal. Assuming the parametrical log-Normal shadowing model for the RSSI signal (defined in equation 4.1) and using the unbiased maximum likelihood (i.e. MLE) distance's estimator $\hat{d} = dK^{-1}10^{\frac{\xi}{10\eta}}$ (defined in equation 4.11), one can deduce that for the considered distances $d_{i,j}$ (or $d_{i,k}$) we obtain their corresponding estimated values $\hat{d}_{i,j}$ (or $\hat{d}_{i,k}$) verifying :

$$\log_{10}\hat{d}_{i,j} = \log_{10} d_{i,j}K^{-1} + \frac{\xi_{i,j}}{10\eta} \quad \rightarrow \quad \log_{10}\hat{d}_{i,j} \sim \mathcal{N}(\log_{10} d_{i,j}K^{-1}, \frac{\sigma^2}{10\eta})$$

From above expression, note that the estimated values of the form $\log_{10}\hat{d}_{i,j}$ are related to the ML estimator of its corresponding true distance $d_{i,j} = |z_i - z_j|$ throughout the logarithm to base 10. Then, we define for each sensor node i the estimated values as $L_{i,j} = \log_{10}\hat{d}_{i,j}$ (respectively for $L_{i,k} = \log_{10}\hat{d}_{i,k}$) which correspond to random variables l following a Normal distribution $\mathcal{N}(\log_{10} dK^{-1}, \frac{\sigma^2}{10\eta})$ whose probability density function (pdf) is written as:

$$f(l) = \frac{1}{2\pi\sigma^2} e^{-\frac{10\eta}{2\sigma^2}(l - \log_{10} dK^{-1})^2}$$

The goal for each sensor node i is to compute its estimated position $\hat{z}_i$ and those of its neighbors $\{\hat{z}_j\}_{\forall j\in\mathcal{N}_i}$ from measured values $\{L_{i,j}\}_{\forall j\in\mathcal{N}_i}$ and $\{L_{i,k}\}_{\forall k\in\mathcal{A}_i}$ of received RSSI packets which depend on distances $\{\hat{d}_{i,j}\}_{\forall j\in\mathcal{N}_i}$ and $\{\hat{d}_{i,k}\}_{\forall k\in\mathcal{A}_i}$ (thus z_i and z_j). In such context, assuming the above parametrical model for the obtained measures at sensor node i we compute the unknown positions using the maximum likelihood criteria. Given a measured value $L_{i,j}$ and its relation $d_{i,j} = |z_i - z_j|$, the maximum likelihood criteria applies as follows :

$$\max_{z_i,z_j} \frac{1}{2\pi\sigma^2} e^{-\frac{10\eta}{2\sigma^2}(L_{i,j} - \log_{10} d_{i,j}K^{-1})^2} \quad \Longleftrightarrow \quad \min_{z_i,z_j} (L_{i,j} - \log_{10}|z_i - z_j|K^{-1})^2$$

We then apply the above optimization criteria to find the set of unknown positions $\{z_i\}_{i=1,\dots,N}$ from the set of measured values $\{L_{i,j}\}_{\forall j \in \mathcal{N}_i}$ and $\{L_{i,k}\}_{\forall k \in \mathcal{A}_i}$ obtained at each sensor node i. Because positions z_i are written under two coordinates (x_i, y_i), we define the vector $\boldsymbol{x} = (x_1, \dots, x_N)$ as the set of the N x-coordinates and in the same way, the vector $\boldsymbol{y} = (y_1, \dots, y_N)$ for the y-coordinates. The global maximum likelihood optimization problem is summarized as:

$$\{\hat{z}_i\}_{\forall i} \;=\; \min_{(\boldsymbol{x},\boldsymbol{y})} \sum_{i=1}^{N} f_i(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)})$$

where

$$f_i(\boldsymbol{x}^{(i)}, \boldsymbol{y}^{(i)}) = \sum_{j \in \mathcal{N}_i} \left(L_{i,j} - \log_{10} |z_i - z_j| K^{-1}\right)^2 + \sum_{k \in \mathcal{A}_i} \left(L_{i,k} - \log_{10} |z_i - \zeta_k| K^{-1}\right)^2$$

In this context, we note that the function to minimize f_i is differentiable as long as we do not consider the case of singular points $z_i = z_j$ and is not convex. When implementing stochastic approximation approach in both centralized ([171]) and distributed ([199]) ways, the convergence results can not be applied directly. As it has been already pointed in [30] among others, the solution is issued to local minima if any random initialization is considered. To avoid this problem and ensure the optimum is achieved, an initialization being a noisy version of the true positions has to be expected. Then, it seems justified the use of such optimization algorithms to improve the inaccurate estimated positions obtained from the already mentioned localization algorithms as multilateration or MDS. A stochastic approximation approach is chosen thanks to its simple implementation by an iterative gradient descent step whose computational design is minimal (i.e. usage of sum and product). However, when the dimension of data under consideration is high (N in our case), the computational and the communication cost may not be suitable for a wireless sensor network with a central node processor. We deal with this problem proposing a distributed version which entails : a local iterative gradient descent step carried out at each individual node to reduce the computational cost. Then a gossip communication step involving two uniformly random selected sensor nodes (from the N possible) at each iteration time to exchange and average their common estimated positions, thus reducing the overall communication cost of the network. Although the centralized algorithm achieves a given accuracy faster than its distributed version, we conclude that a reasonable and well-suited trade off is obtained for wireless sensor network technology.

The distributed stochastic algorithm proposed to solve the previous minimization problem involves two steps at each iteration time t: each sensor node i maintains an estimation of $\boldsymbol{z}_t^{(i)}$ at time t and performs a local update of its positionÕs set based on a classical gradient descent (GD) step. And later all sensor nodes exchange their estimated $\boldsymbol{z}_t^{(i)}$ with its neighboring sensor node. This last step of the algorithm works in two steps at each iteration time t:

- **Local gradient descent step [171]**

  Each sensor i generates a temporary estimate of its known positions' set based on the local current RSSI measurements. Those are the positions of the sensor nodes from which it receives packets and its own position:

  $$(\tilde{\boldsymbol{x}}_{t+1}^{(i)}, \tilde{\boldsymbol{y}}_{t+1}^{(i)}) = (\boldsymbol{x}_t^{(i)}, \boldsymbol{y}_t^{(i)}) - \gamma_t \nabla f_i(\boldsymbol{x}_t^{(i)}, \boldsymbol{y}_t^{(i)})$$

  where $(\gamma_t)_{t \geq 1}$ is a decreasing step: $\gamma_t = 1/\sqrt{t}$.

- **Gossip step [33]**

  A sensor node i is uniformly randomly selected, then it chooses randomly a node j in its neighborhood $\mathcal{N}_i$ for the position consensus phase. For this step, we use a communication protocol known as Pairwise in [33]. Each sensor node i from the set of N sensor nodes, estimates its own position $(x_{t+1}^{(i)}(i), y_{t+1}^{(i)}(i))$ as well as those of its neighboring sensor nodes $\{(x_{t+1}^{(i)}(j), y_{t+1}^{(i)}(j))\}_{\forall j \in \mathcal{N}_i}$.

  Let $\mathcal{N}_i$ the set of neighboring sensor nodes i and $\mathcal{N}_j$ the set of neighboring sensor nodes of sensor j. During this position exchange step, each sensor node randomly selects the neighbor with which it will communicate its list of position estimates; and therefore both neighboring sensor nodes mutually exchange their respective position but also those estimated for their respective neighboring positions. It should be noted that during this position exchange process, only the positions of neighbors lying in both of the two sets $\mathcal{N}_i$ and $\mathcal{N}_j$ will be updated by the two neighboring sensor nodes i and j. Estimated sensor nodes positions from nodes that will not be in both of the two sets $\mathcal{N}_i$ and $\mathcal{N}_j$ will not be shared by the two neighbors i and j. For example, consider two sensor nodes: the sensor node 1 has estimated its position (x_1, y_1) and those of its neighbors $\mathcal{N}_1 = \{2, 3, 4\}$ while the sensor node 5 has estimated its position (x_5, y_5) and those of its neighbors $\mathcal{N}_5 = \{2, 3, 7, 8\}$. Each of the sensor nodes 1 and 5 exchange their estimated positions and those from the shared nodes 2 and 3. The position estimation of the other neighbors: 4 for the sensor node 1, 7 and 8 for the sensor node 5 will not be exchanged in this iteration in which the sensor nodes 1 and 5 communicate.

  The two uniformly randomly selected sensor nodes involved in the Gossip step merge their common estimated positions and average their values as follows:

$$\forall \ell \in \mathcal{N}_i \cap \mathcal{N}_j, \tag{5.10}$$

$$\begin{aligned}(\boldsymbol{x}_{t+1}^{(i)}(\ell), \boldsymbol{y}_{t+1}^{(i)}(\ell)) &= \frac{(\tilde{\boldsymbol{x}}_{t+1}^{(i)}(\ell), \tilde{\boldsymbol{y}}_{t+1}^{(i)}(\ell))}{2} + \frac{(\tilde{\boldsymbol{x}}_{t+1}^{(j)}(\ell), \tilde{\boldsymbol{y}}_{t+1}^{(j)}(\ell))}{2} \\ (\boldsymbol{x}_{t+1}^{(j)}(\ell), \boldsymbol{y}_{t+1}^{(j)}(\ell)) &= (\boldsymbol{x}_{t+1}^{(i)}(\ell), \boldsymbol{y}_{t+1}^{(i)}(\ell))\end{aligned} \tag{5.11}$$

  In other cases, the rest of sensor nodes not concerned in the communication phase maintain their temporary estimated positions as: $\forall \ell \notin \mathcal{N}_i \cap \mathcal{N}_j$ and $\forall m \neq i, j$ then $(\boldsymbol{x}_{t+1}^{(m)}(\ell), \boldsymbol{y}_{t+1}^{(m)}(\ell)) = (\tilde{\boldsymbol{x}}_{t+1}^{(m)}(\ell), \tilde{\boldsymbol{y}}_{t+1}^{(m)}(\ell))$.

To present the results of this method and compare them to the previous ones, we analyze its behavior in a network of $N = 16$ sensor nodes in SensLAB Rennes platform whose configuration is shown in Figure 4.4. First Data are simulated using Monte Carlo with additive noise using different noise levels values σ^2. Initialization points are randomly chosen for each sensor node according to two uniform components between 0 and 5. We illustrated in Figure 5.16, the normalized Mean Error named NME (cf. equation (5.9) of N positions using 1000 independent Monte Carlo realizations (RSSI data generated with Matlab)) with 1000 iterations and for each value of σ^2 ranging from 0 and 10.

Figure 5.17 shows the decreasing mean error through the time (as the number of iteration t increases) in the simulated data case. Note that the index of iterations is a discrete time index. For example, it is worth watching the value taken by the sequence (γ_t) used by the iterative

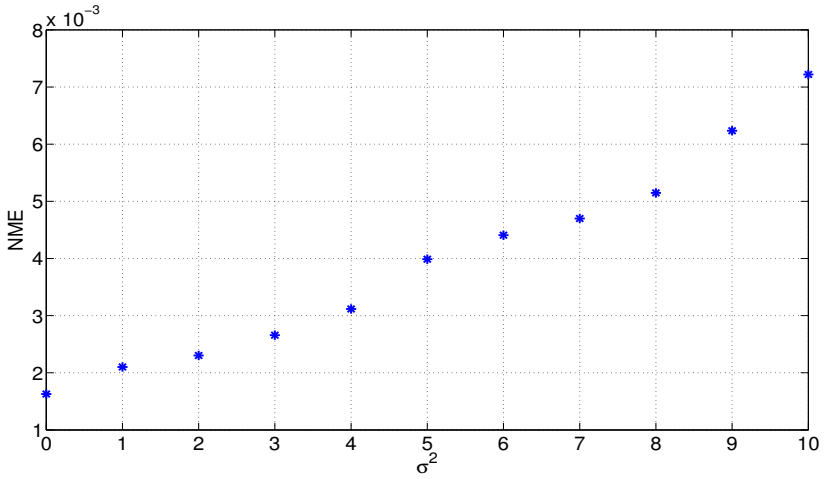


Figure 5.16: Normalized mean error of $N = 12$ unknown positions estimated from noisy data simulated with log-normal noise (different values of σ^2) and using the Distributed Stochastic Approximation (DSAL) algorithm. We consider a random initialization and a decreasing step of $\gamma_t = 1/\sqrt{t}$.

algorithm DSAL when it is executed at iteration of discrete time t with $t \geq 1$. $t \geq 1$ thus defines the discrete time index when the DSAL algorithm starts working and generates the estimated sequences $(\boldsymbol{x}_t)$ and $(\boldsymbol{y}_t)$. The DSAL algorithm is applied using constant and decreasing step size and data generated with Matlab from the noisy estimated distances as the log-normal model (cf. equation 4.11). The data is drawn when we set the parameters $T = 1$ number of RSSI measure, $\eta = 2.1$ and the noise variance $\sigma^2 = 16.37$.

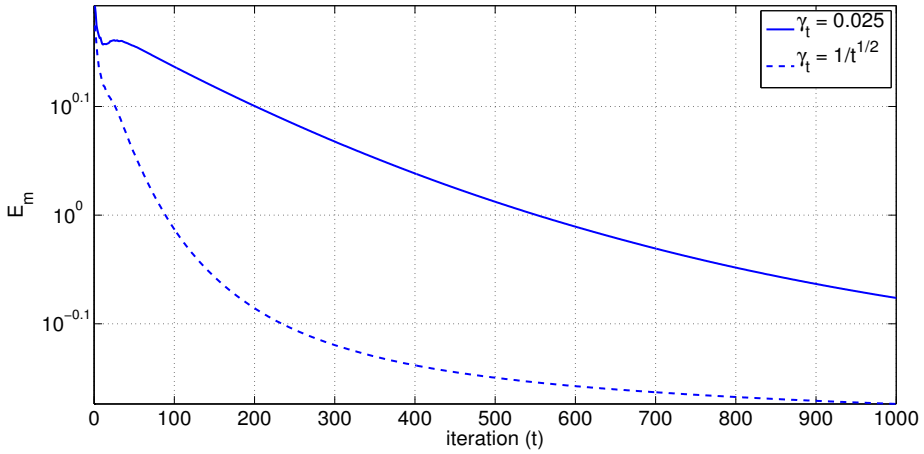
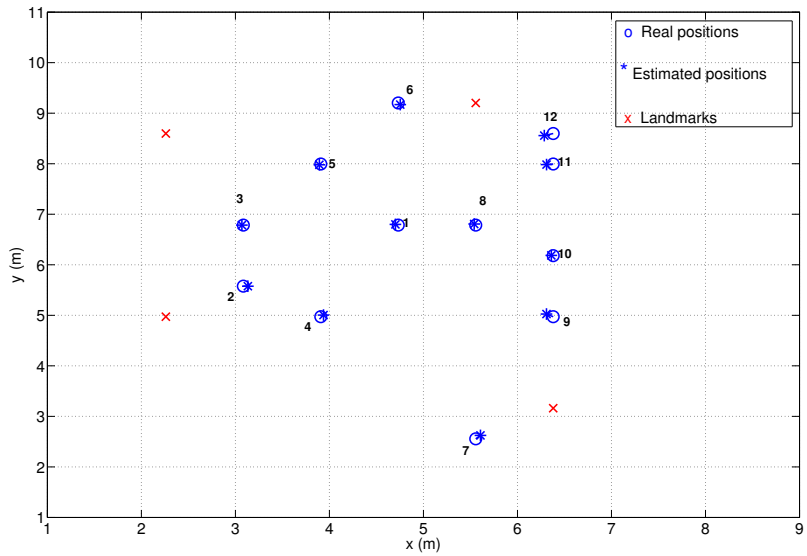


Figure 5.17: Mean error (see equation (5.8)) of the $N = 12$ unknown positions estimated at each iteration t while using the DSAL algorithm. A constant step size $(\gamma_t = 0.025)_{t\geq 1}$ and decreasing step size $(\gamma_t = 1/\sqrt{t})_{t\geq 1}$ are considered. Simulated data with Matlab is used to draw noisy estimated distances as (cf. equation4.11) with $T = 1$, $\eta = 2.1$ and log-normal noise with variance $\sigma^2 = 16.37$.

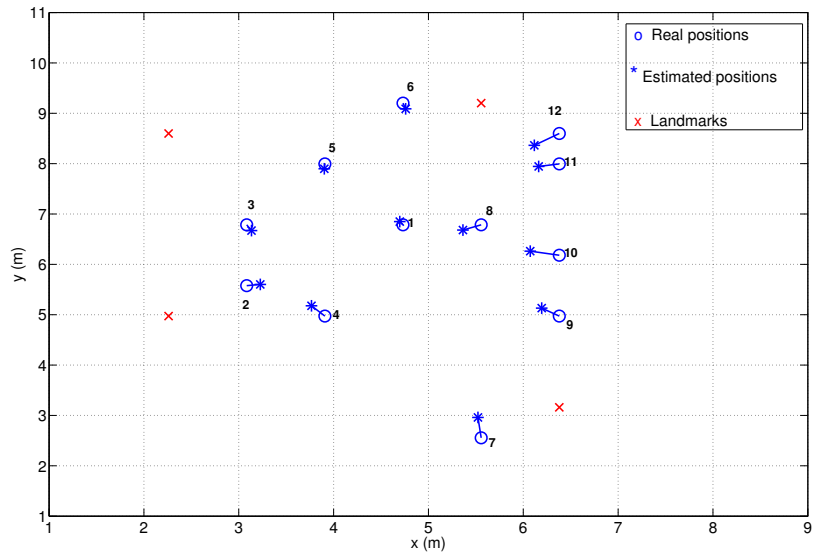
Figures 5.18(a) and 5.18(b) show the average results of the estimated positions using the Distributed Stochastic Approximation algorithm (named DSAL) when σ^2 is equal to 2 and 10.

While Figures 5.19(a) and 5.19(b) illustrate the result when one data is generation with Matlab.

After validating the algorithm in simulation, we will now experiment its results using real data coming from remote large scale wireless sensor nodes in SensLAB Rennes platform. 10 indepen-

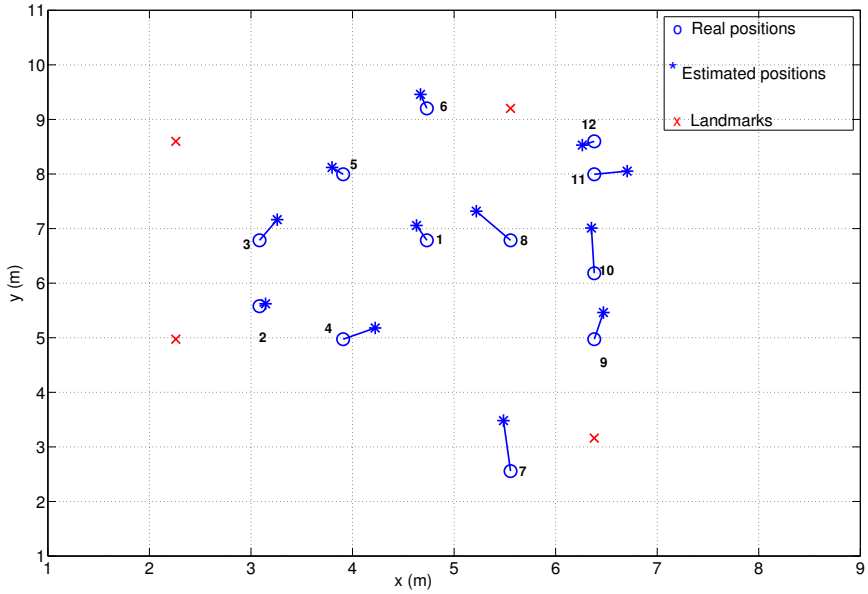


(a) Noise variance: $\sigma^2 = 2$.

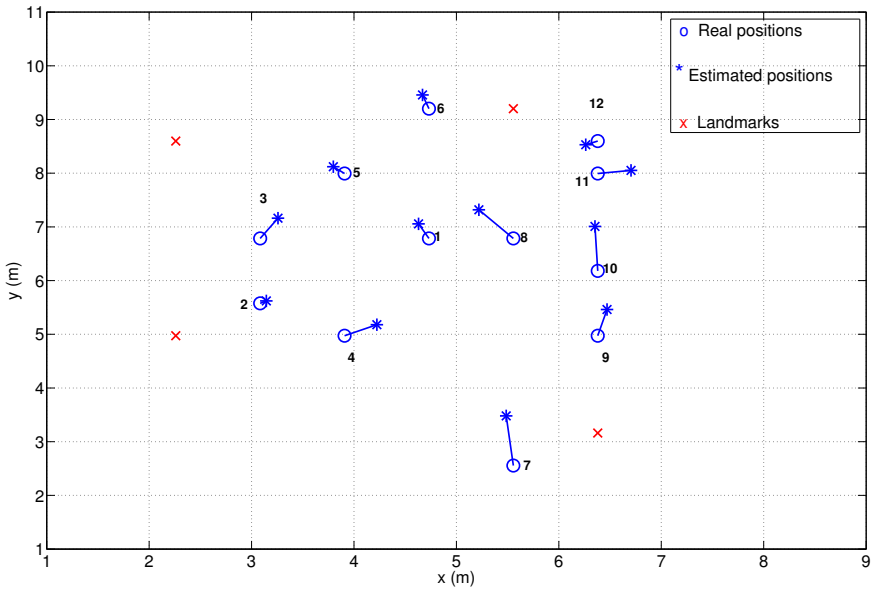


(b) Noise variance: $\sigma^2 = 10$.

Figure 5.18: Average of the $N = 12$ unknown positions estimated with Distributed Stochastic Approximation (DSAL) algorithm using 1000 independent Monte Carlo realizations (RSSI data generated with Matlab)



(a) Noise variance: $\sigma^2 = 2$.



(b) Noise variance: $\sigma^2 = 10$.

Figure 5.19: $N = 12$ unknown positions estimated with the Distributed Stochastic Approximation algorithm (DSAL) with the use of one independent Monte Carlo realization (one round of RSSI data generated with Matlab)

dent experiments are realized in the SensLAB Rennes platform. In each experiment, each sensor node collects until 600 packets. We select 12 blind nodes (including virtual landmarks) and 4 landmarks. The average positioning estimates with a random guess/initialization and after $t = 1000$ iterations are presented in Figures 5.20(a) and 5.20(b). We consider a decreasing step value of $(\gamma_t)_{t\geq 1}$ at each iteration as $1/\sqrt{t}$. The DSAL runtime for the estimation of 12 unknown nodes positions and after $t = 1000$ iterations is 15.16 seconds. Furthermore, the normalized mean error (NME) of N unknown positions estimated under 10 independent experiments is 0.14. The average error value for each position of each sensor node is:

$$(0.1, 0.21, 0.09, 0.15, 0.1, 0.05, 0.11, 0.16, 0.47, 0.22, 0.11, 0.08).$$

The same algorithm is evaluated with the same error estimation under 100 independent experiments (60 RSSI values are measured by each sensor node). A normalized mean error (NME) of 0.135 is obtained. For each sensor, the position estimation is:

$$(0.1, 0.21, 0.08, 0.16, 0.09, 0.05, 0.1, 0.15, 0.51, 0.18, 0.1, 0.07).$$

Figures 5.21(a) and 5.21(b) show the mean estimation of the $N = 12$ positions after 1000 iterations. We consider a randomly initialization of the DSAL algorithm and a constant local gradient step through each iteration; we set $\gamma_t = 0.025$ for all t. In such case, the normalized mean error (defined in equation 5.9) of N positions from 10 independent experiments is 0.082. The estimated normalized mean error (NME) for each sensor node is summarized as:

$$(0.05, 0.11, 0.03, 0.07, 0.041, 0.04, 0.07, 0.09, 0.35, 0.09, 0.09, 0.06).$$

When running 100 independent experiments (60 packets received at each sensor node at the SensLAB platform) the NME value is 0.083. The individual NME estimation at each sensor node is:

$$(0.05, 0.1, 0.06, 0.09, 0.06, 0.05, 0.07, 0.08, 0.34, 0.09, 0.06, 0.07).$$

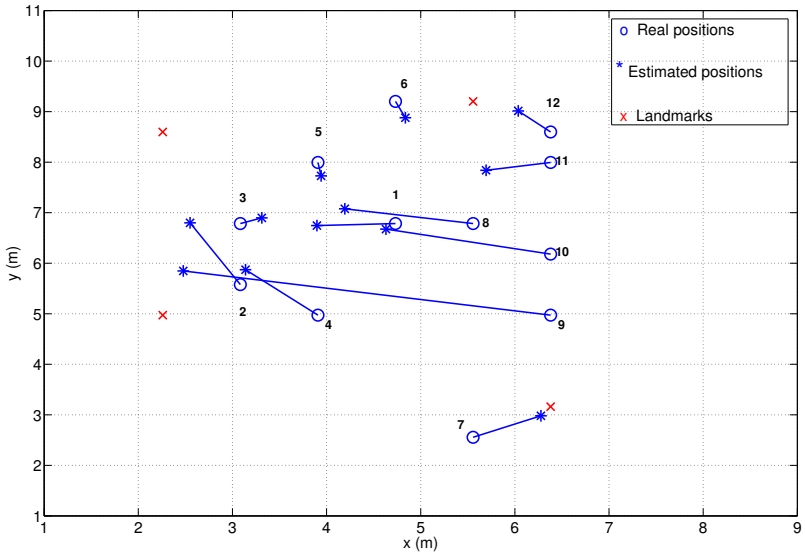
In Figures 5.22(a) and 5.22(b) we show the estimation of $N = 12$ sensor node positions from 10 independent experiments at SensLAB platform and using the DSAL algorithm with constant and decreasing step size respectively. In each experiment, the distributed algorithm is initialized with the MLE position estimator from the RSSI data received from the landmarks at each sensor node (using the *fminsearch* function provided in Matlab). We computed the NME error from the 10 experiments after 1000 iterations.

If we take a constant step size $\gamma_t = 0.025$ for all iteration t, the NME is 0.076 and the mean error at each sensor node position is:

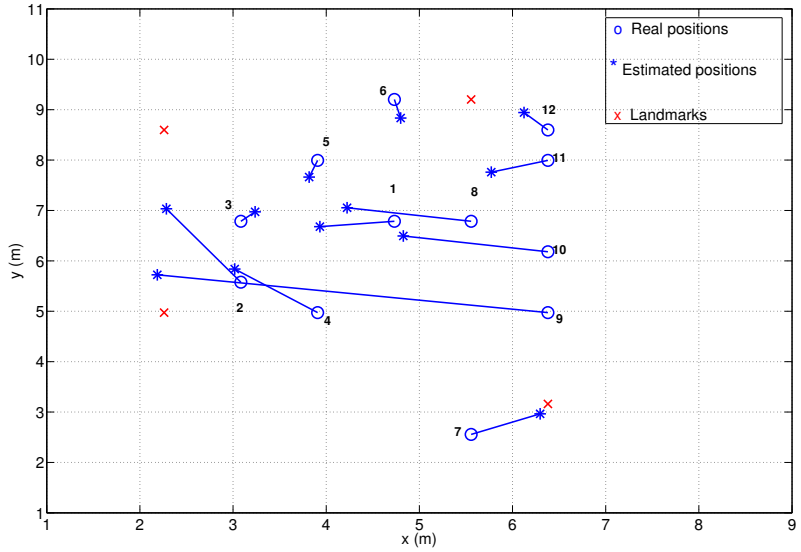
$$(0.02, 0.064, 0.067, 0.048, 0.066, 0.03, 0.06, 0.088, 0.35, 0.07, 0.032, 0.012).$$

Meanwhile taking the decreasing step size as $(\gamma_t = 1/\sqrt{t})_{t\geq 1}$, the NME value is slightly higher 0.1 in opposite as we expected from the simulated data model (cf. Figure 5.17). The mean error at each sensor node position is:

$$(0.041, 0.095, 0.05, 0.073, 0.049, 0.075, 0.086, 0.1, 0.43, 0.13, 0.044, 0.065)$$

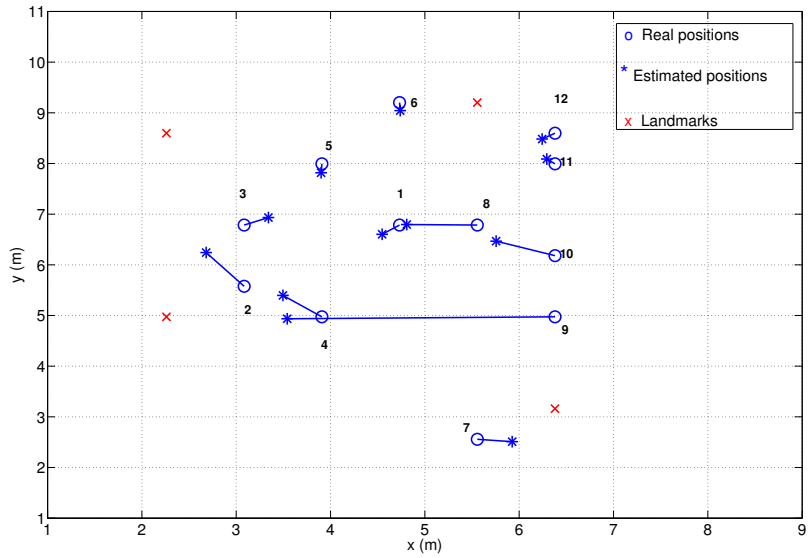


(a) Mean estimation of $N = 12$ positions under 10 independent experiments in the SensLAB Rennes platform.

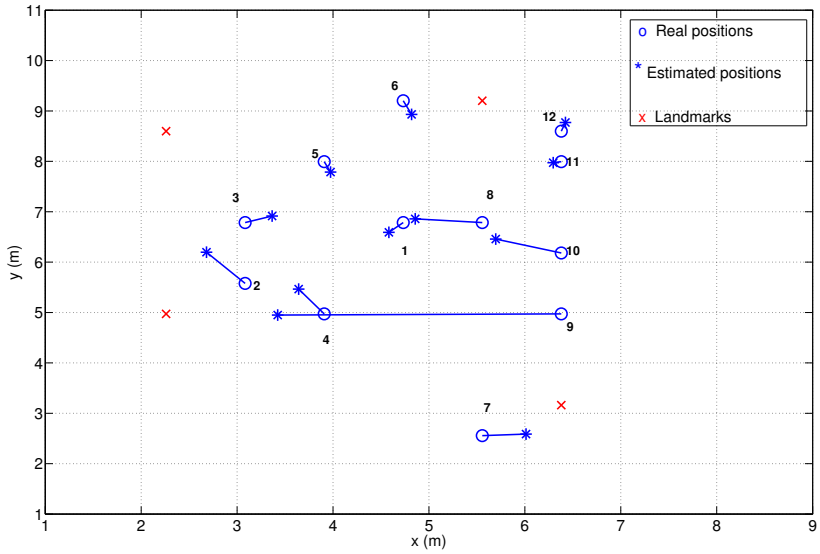


(b) Mean estimation of $N = 12$ positions under 100 independent experiments in the SensLAB Rennes platform.

Figure 5.20: Mean estimation of $N = 12$ positions with the Distributed Stochastic Approximation algorithm (DSAL) with random initialization under 10 independent experiments in the SensLAB Rennes platform. A decreasing step value of $(\gamma_t = 1/\sqrt{t})_{t\geq 1}$ is considered

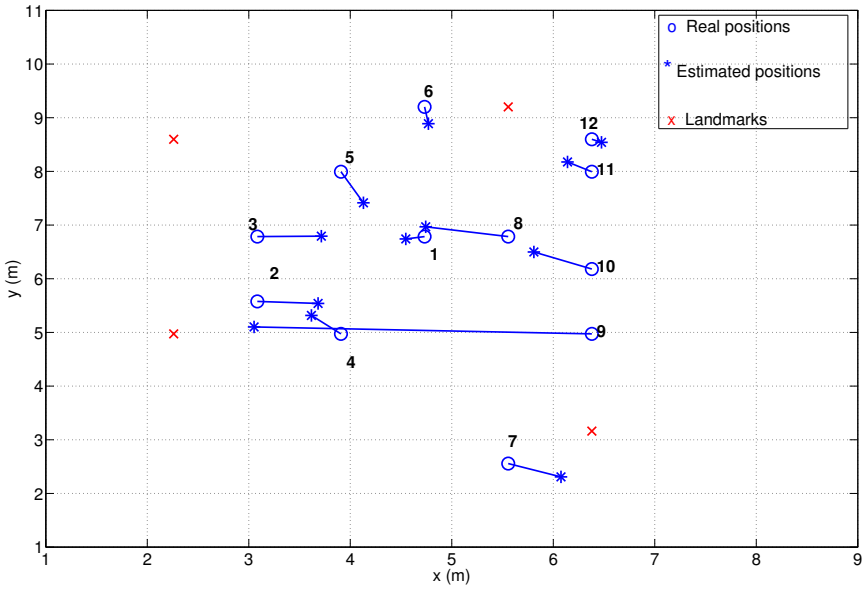


(a) Average positions estimation from 10 independent experiments.

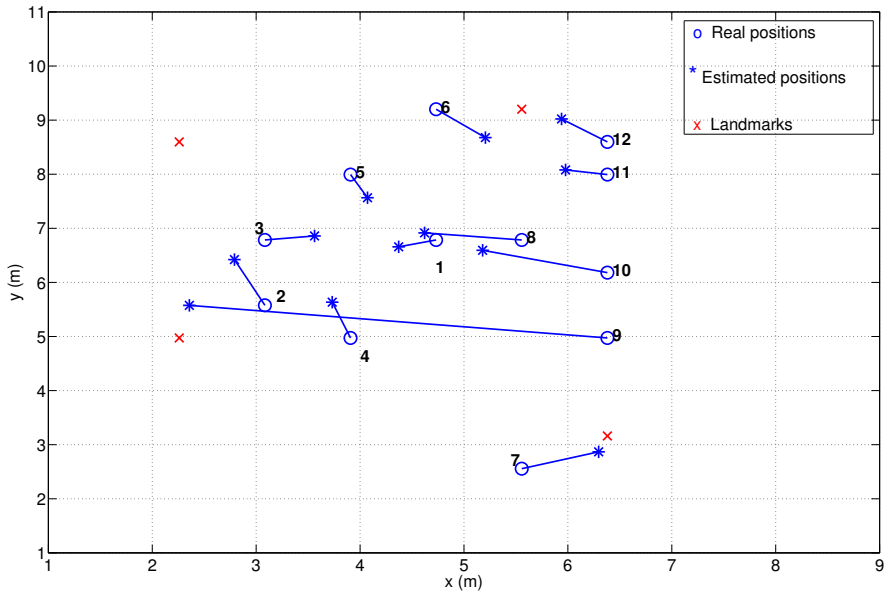


(b) Average positions estimation from 100 independent experiments.

Figure 5.21: Average of the resulting $N = 12$ unknown positions estimated with the DSAL algorithm with random initialization and a constant step size $(\gamma_t = 0.025)_{t\geq 1}$. Experiments are realized at SensLAB Rennes platform.



(a) Constant step size $(\gamma_t = 0.025)_{t\geq 1}$.



(b) Decreasing step size $(\gamma_t = 1/\sqrt{t})_{t\geq 1}$.

Figure 5.22: Average of the resulting $N = 12$ unknown positions estimated from SensLAB experiments when using the DSAL algorithm with MLE initialization. Both, constant and decreasing step size are considered.

5.3 Summary and discussion on distributed algorithms results

In order to compare the proposed method to the several ones presented above, we set the sensor node configuration with N equal to 16 as shown in Figure 4.4. In such network, we consider 12 blind sensor nodes including virtual landmarks and 4 landmarks. We analyze the resulting mean error (defined in equations (5.9) and 5.8) from the simulated data in Matlab first and from the real data collected at SensLAB Rennes platform then. We compare the performance of the following three methods: multilateration (named MC), multidimensional scaling (classical MDS and SMACOF) and distributed stochastic approximation (DSAL) also named Distributed Gradient (DG). We show in Figure 5.23 the four curves of the normalized mean error (NME) of the 12 unknown positions estimated under 1000 independent Monte Carlo simulations run of each algorithm. We vary the variance noise parameter σ^2 from 0 to 20. The simulated data with Matlab is drawn as noisy estimated distances from the log-normal model (unbiased estimators in equations (4.11) and (4.13) when $T = 1$ RSSI reading and $\eta = 2.1$). This value for the parameter η corresponds to the mean estimated from RSSI measurements from the SensLAB nodes.

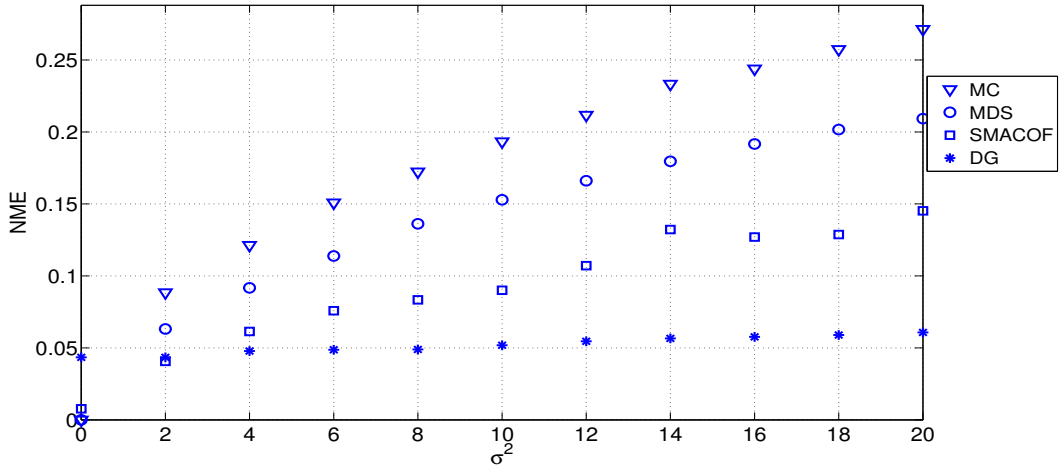


Figure 5.23: Normalized mean error of $N = 12$ unknown positions estimated with simulated data from 1000 Monte Carlo trials depending on σ^2. With DSAL algorithm, a constant step size $(\gamma_t = 0.025)_{t\geq 1}$ and a random initialization is considered

When using the real data from the SensLAB experiment, a first phase involves the estimation of the parameters PL(d_0), η and σ^2 (see Table 4.1); if we assume the log-normal shadowing model of the RSSI readings from the packets collected at the SensLAB nodes platform. We summarized in Table 5.1 the values of the normalized mean error (NME) and the mean error in meters (E_m) for the three methods recalled above and for the MLE estimation. Both mean errors are computed from 10 independent experiments in which each sensor node collected until 600 packets.

Table 5.2 details the NME for each of the 12 unknown sensor nodes positions estimated with the same different methods described above and using real SensLAB data. DSAL results are obtained with a constant step size $(\gamma_t = 0.025)_{t\geq 1}$ and MLE location initialization. $\eta = 2.1$, $\sigma^2 = 16.37$ and $T = 1$ RSSI measure.

Finally, Figure 5.24 shows the normalized mean error (NME) of $N = 12$ unknown positions estimation and when using SensLAB data. Positioning results for the selected network illustrated in Figure 4.4 are summarized in Table 5.2.

Table 5.1: Normalized and absolute mean error values (NME and E_m) estimated for 12 sensor nodes positions from SensLAB real data using different methods explained in above sections

Method	**NME (%)**	$\boldsymbol{E_m}$ **(m)**
Multilateration (MC)	0.23	2.2
Classical MDS	0.3	2.8
Modern MDS (SMACOF)	0.25	2.4
MLE	0.28	2.63
DG (decreasing step and random ini.)	0.14	1.3
DG (constant step and random ini.)	0.08	0.77
DG (decreasing step and MLE ini.)	0.1	0.98
DG (constant step and MLE ini.)	0.076	0.72

Table 5.2: Values of the NME for each of the 12 unknown positions to be estimated using different methods when collecting data from 10 independent experiments at SensLAB platform.

	Simulated Matlab data				**Real SensLAB data**			
NodeID	**MC**	**MDS**	**SMACOF**	**DG**	**MC**	**MDS**	**SMACOF**	**DG**
214 / 1	0.15	0.13	0.093	0.051	0.13	0.23	0.19	0.02
176 / 2	0.21	0.17	0.13	0.056	0.2	0.29	0.17	0.064
178 / 3	0.20	0.16	0.13	0.064	0.05	0.05	0.27	0.067
193 / 4	0.19	0.15	0.12	0.056	0.21	0.35	0.23	0.048
198 / 5	0.23	0.16	0.11	0.053	0.2	0.02	0.15	0.066
218 / 6	0.34	0.22	0.11	0.054	0.45	0.14	0.28	0.03
225 / 7	0.41	0.27	0.15	0.071	0.22	0.59	0.51	0.06
232 / 8	0.17	0.16	0.1	0.051	0.21	0.12	0.24	0.088
247 / 9	0.23	0.21	0.13	0.06	0.2	0.47	0.063	0.35
249 / 10	0.21	0.20	0.12	0.054	0.25	0.29	0.41	0.07
252 / 11	0.28	0.23	0.15	0.052	0.25	0.32	0.35	0.032
253 / 12	0.33	0.25	0.15	0.055	0.35	0.12	0.18	0.012

Regarding the achieved results, localization with multilateration fails when the sensor nodes to localize lies at the outside bound of the network. For instance, those nodes with identifier 218 and 253 have higher mean errors with multilateration. In the same way, the worst localized node positions with MDS belong to nodes with identifier 225 and 247 which are at outside the bound of the landmarks. However with the DSAL algorithm, we have the lowest mean error for all sensor nodes unless for those with identifier 247 and 193 in which SMACOF and MDS algorithms respectively outperform. Note that the sensor node with identifier 247 has the lowest η parameter which value is 0.66 and high noise variance σ^2 which value is 14.09. From our experiments and simulation results, we have shown the proposed DSAL distributed localization algorithm outperforms all the other compared algorithms. Indeed, it can achieve a localization error of 0.77 m in a confined area of 5 m by 8 m with four landmarks. One should note that, when using the refinement step of DSAL algorithm, blind node and virtual landmarks nodes positions can be estimated at the same time with the DSAL with at least two landmarks. However, in order to make a comparison with multilateration approach, four landmarks are used with DSAL.

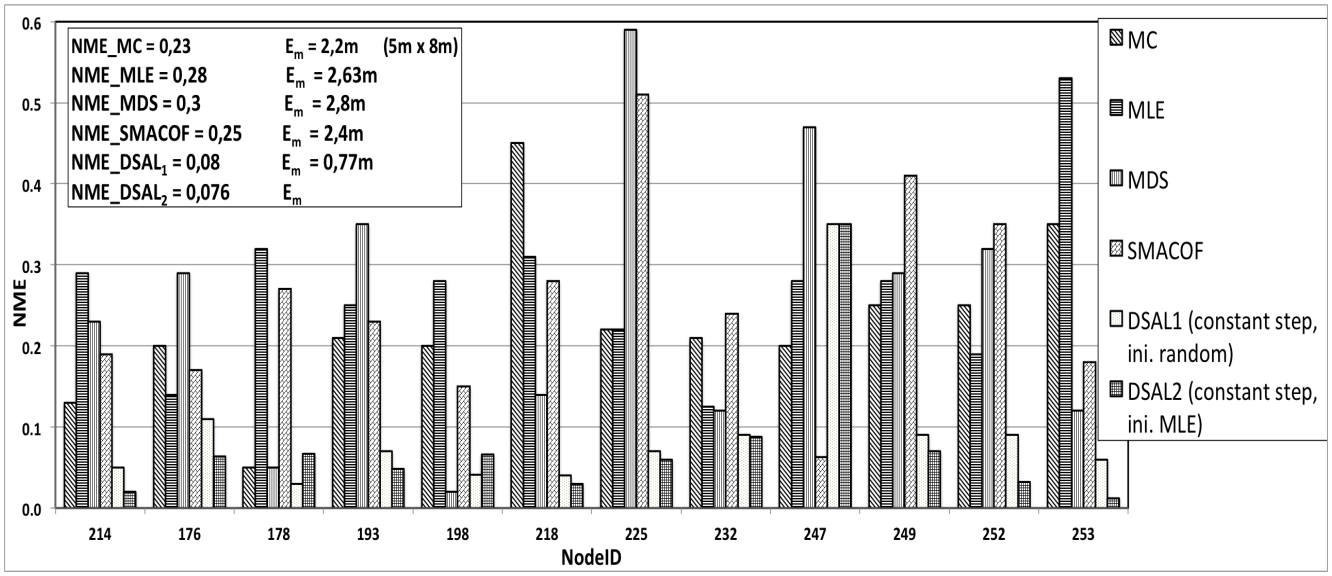
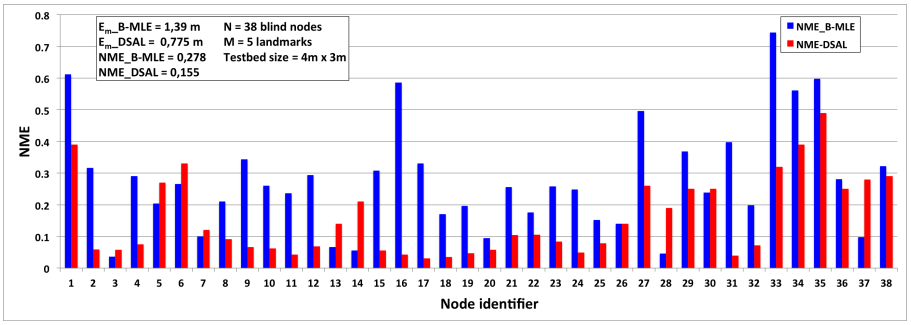


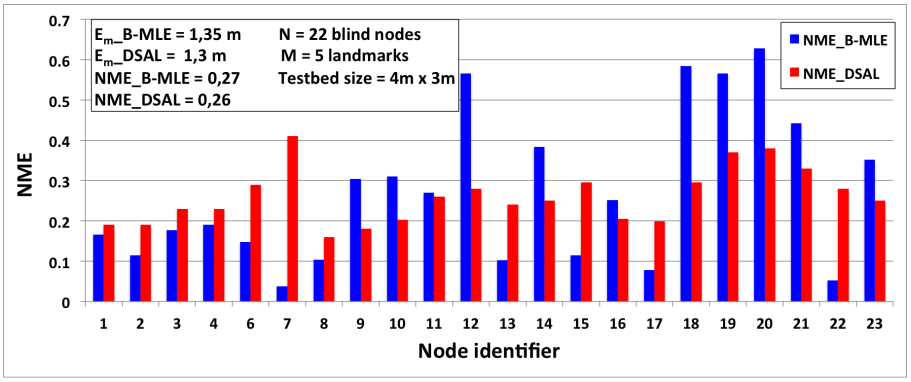
Figure 5.24: Normalized mean error (NME) for each unknown sensor node position estimated with different methods using the real data collected from 10 independent experiments runs at the SensLAB platform. The network has 12 blind sensor nodes (unknown positions to be estimated including virtual landmarks) and 4 landmarks.

DSAL is also able to improve the accuracy of blind nodes locations initially estimated by our proposed B-MLE algorithm (cf. Figures 5.25 and 5.26). At Lincs testbed, localization errors of 0.775 m and 1.3 m are obtained by applying DSAL algorithm on B-MLE localization results which are respectively 1.39 m and 1.35 m for the small (*i.e.* 10 positions) and large (*i.e.* 25 positions) learning set. At Rammus testbed (Rennes platform), DSAL refines the location accuracy at about 1.73 m and 1.64 m respectively while B-MLE obtains a mean error of 2.73 m and 2.20 m for a small (*i.e.* 10 positions) and large (*i.e.* 25 positions) learning set.

Finally, we notice that for all blind nodes positions to localize at the Lincs testbed, DSAL has refined localization accuracy at 79% and 55% compared to B-MLE respectively for the small (*i.e.* 10 positions) and large (*i.e.* 25 positions) learning set. The accuracy is 72% and 74% on the Rammus testbed with respect to B-MLE for both, the small (*i.e.* 10 positions) and the large (*i.e.* 25 positions) learning set (cf. Table 5.3).

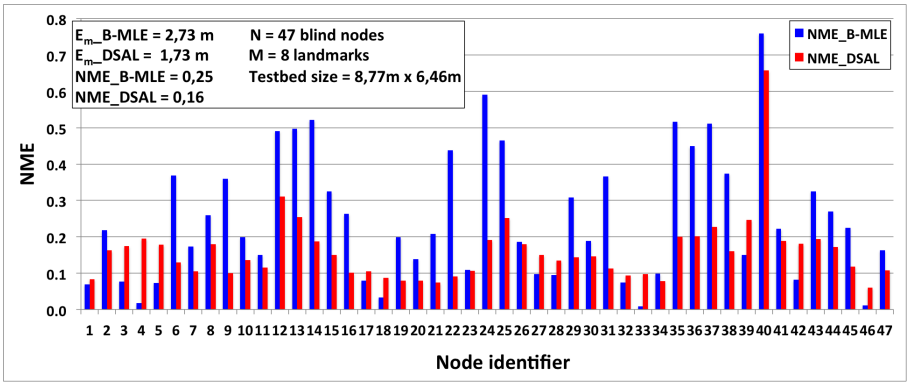


(a) DSAL localization performance with a decreasing step ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) and B-MLE initialization. Results from Lincs testbed when considering a small learning set size (10 positions).

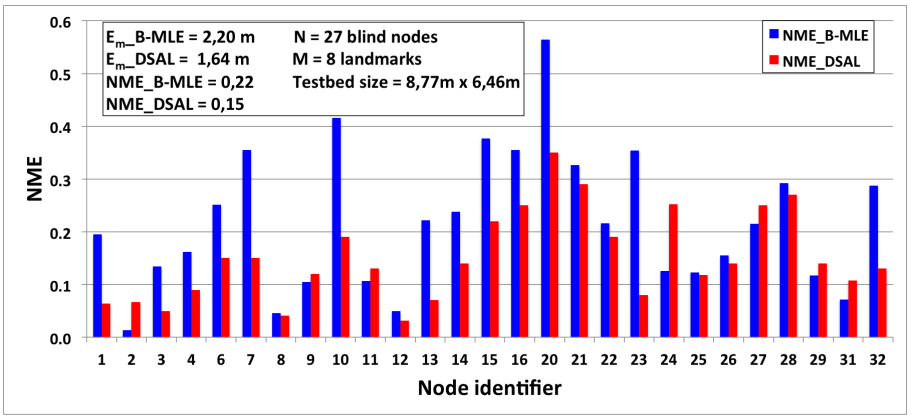


(b) DSAL localization performance with a decreasing step ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) and B-MLE initialization. Results from Lincs testbed when considering a big learning set size (25 positions).

Figure 5.25: DSAL localization performance with a decreasing step ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) and B-MLE initialization. Results from Lincs testbed and two different learning set size.



(a) DSAL localization performance with a decreasing step ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) and B-MLE initialization. Results from Rammus testbed when considering a small learning set size (10 positions).



(b) DSAL localization performance with a decreasing step ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) and B-MLE initialization. Results from Rammus testbed when considering a big learning set size (10 positions).

Figure 5.26: DSAL localization performance with a decreasing step ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) and B-MLE initialization. Results from Rammus testbed and two different learning set size.

Table 5.3: DSAL localization results with B-MLE initialization (our previous proposal) and a decreasing step sequence such $(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$

Characteristics	**Lincs testbed-small learning set**	**Lincs testbed-bigger learning set**	**Rammus testbed-small learning set**	**Rammus testbed-bigger learning set**
Number of blind nodes positions	38	22	47	27
Number of positions with whom localization Error increases w.r.t DSAL	8	10	13	7
Number of positions with whom localization Error decreases w.r.t DSAL	30	12	34	20
Percentage of position well estimated w.r.t DSAL compared to B-MLE	79%	55%	72%	74%

Chapter 6

Conclusions and suggestions for further work

Contents

6.1 Summary

*** Le début ne laisse pas présager la fin. (—Hérodote—) ***

This thesis has investigated the problem of indoor localization using RSSI (i.e. Received Signal Strength Indicator) measurements. Localization in indoor environment is a challenging issue mainly due to both constraints: first the unavailability of global positioning system (i.e. GPS) in indoor. GPS signals are attenuated in indoor due to buildings, basements construction materials (concretes walls, metal) that block them. Indeed, in the case where few GPS signals can bypass such obstructing factors, they are not fine accurate (about 10 meters) to be used in indoor environment. GPS receivers are costly in terms of price. Such receivers are also power consuming and thus costly in term of energy consumption to be supported by low power devices like wireless sensor nodes. Such embedded tiny nodes are micro components limited in terms of processing, storage capability and they often have to keep energy consumption low. Second, indoor environment is dynamic. It comprises several obstacles, furniture pieces, materials (glass, wood, bricks, iron materials, roofing tiles, wall, plywood), moving people and objects. Such obstacles are disposed and arranged in confined areas and can reflect, scatter as well as diffract the radio signal causing at the same time multipath and shadowing. Hence, performing localization in such uncontrolled scenario is non trivial and a difficult issue to solve.

Furthermore, the measurement of the attenuation of the radio signal along the path between a radio emitter and receiver named RSSI is the only economically feasible and available measurement in IEEE 802.15.4 devices. The RSSI has been used in most of the localization algorithms in indoor and outdoor for inter-node distance estimation also called ranging. In fact, the RSSI does not require any time synchronization among devices and no extra hardware compared to the time of flight of the radio signal (ToA or ToF) or the use of angle of arrival of the signal (AoA). Both ToA and AoA yield a better accuracy but are more energy demanding and require more expensive equipments and thus are very difficult to measure compared to RSSI. This latter is directly measurable from each supplied wireless sensor node radio chip. However, the unpredictable nature of the indoor environment propagation behavior makes the use of RSSI an inaccurate and feeble technique affected by indoor obstacles. Consequently, on the localization point of view, when considering an adaptable localization algorithm to work in any indoor

scenario, studying the RSSI behavior according to indoor radio signal propagation is needful. In fact, such analysis helps to characterize the RSSI dispersion as well as the widely used and being less complex propagation model: the log-normal shadowing model. To this aim, we have realized several distinct measurements campaigns in different indoor testbeds and in SensLAB, a remote large scale wireless sensor network. These measurements campaigns are performed to identify the behavior of the RSSI in relation to distance, data measurement time, presence or not of people. In fact the wireless signal is absorbed by human body and affected by variation over time. Such situations are examined under different configurations including RSSI behavior in an anechoic chamber, in a large as well as a small room with different emitter $\leftrightarrows$ receiver deployment scenarios in random grids. The antenna gain of the sensor node we use; the Tmote Sky mote has also been characterized and shows a non-omnidirectional pattern. The radiation pattern is not entirely uniform, nor smooth, a few directions experience a strong attenuation due to electronics on one of the board and reflections effects from the wall, furniture. Such effects, even when walls are non-concrete walls are noticeable and are more representative in small room.

Our experiments have also highlighted that the measured RSSI do not decreases regularly with distance as stated in many previous research efforts. We have also showed that the high variation of RSSI cannot be explained only by the log normal shadowing model. We additionally observe that RSSI values are dependent to radio emitter node environmental location with respect to a receiver. Consequently, RSSI readings only match the log-normal shadowing model pattern when they are considered independently according to each emitter node in the localization system. Overall measurements results from our experimental studies and from the several other ones that were published, conclude that the RSSI can sometimes, on some links, provide a good indication on the distance. On the other hand, in many situation, the mapping between signal attenuation and emitter-receiver distance is far from obvious. In a dense network (i.e. when each node possesses strictly more than 3 neighbors) it should be possible to improve localization accuracy by carefully selecting the links that are taken into account for ranging.

Thereby, having an accurate and efficient localization algorithms adaptable to all types of indoor environment such as from office buildings, room, tunnels, basement, mall, better address the problem of localization in these dynamic environments. Such environments are also variable over time and are dependent to the instability of the radio channel propagation. To this ambition, as a second objective of this PhD thesis, we have proposed and evaluated a family of algorithms to enhance indoor positioning of wireless devices using RSSI measurements from data packets coming from multiple sources named landmarks. In addition, such adaptable localization algorithms can be performed with no pre-deployment efforts. We assume a blind mobile node to evolve in an unknown environment composed of some landmarks (minimum of three) with known positions. A person or an object that are carrying this blind node can estimate its own $2D$ coordinates in the environment using RSSI measurements from landmarks while running one of the available localization algorithms. As our proposed localization algorithms use ranging from the RSSI measurements, note that the propagation model parameters are computed online together with the position estimate. Instead, in most previous works, the propagation model parameters are computed offline and kept constant while separated from the localization part.

Thus, in this thesis, as a third objective, we have observed the problem of solving 2D coordinates of any blind mobile node from one-hop then distributed localization algorithms point of view. We based our study on real sensor nodes data in real environments. Uncontrolled scenarios show that if some sources are coherent, some are clearly aberrant and introduce high shadowing variance and multipath. Therefore, after studied the localization results of RSSI-based maximum likelihood estimator (i.e. MLE) compared to a trilateration and a random localization mecha-

nism, we have proposed and evaluate a family of algorithms to enhance indoor positioning of wireless devices using RSSI measurements from data packets coming from multiple landmarks. We first proposed a dynamic method to identify an abnormal landmark and to reduce its effects, replacing its measurements by a constant bias. This dynamic method named Biased-maximum likelihood estimator (i.e. B-MLE) only uses the necessary number of active landmarks. This algorithm is evaluated in real testbeds with different configurations. Experimental results have demonstrated B-MLE to achieve a more accurate result than classical localization algorithms like maximum likelihood applied on a classical log-normal shadowing model. We then envisaged a general heuristic localization mechanism that directly reduces the effects of abnormal landmarks based on a threshold on the bias and thus does not need to examine all possible subsets of landmarks. This generic heuristic localization mechanism is called Biased-Maximum Likelihood Location Estimation extension Correction (i.e. B-MLEC). Compared to the first approach (i.e. B-MLE), the generic heuristic localization mechanism (i.e. B-MLEC) is more efficient in smaller testbeds which seem to be more affected by indoor obstacles than larger testbeds in which ignoring a single abnormal landmark is the best method. But both have ameliorated the RSSI-based indoor location estimation compared to MLE while demonstrating a better estimation of the propagation model parameters.

With an objective to offer a better location accuracy, we have also proposed another efficient RSSI-based indoor localization algorithm named GMM-MLE for Gaussian Mixture Model. This GMM-MLE simple heuristic localization algorithm is based on statistical machine learning approach. GMM-MLE detects and filters aberrant patterns while identifying the distribution of acquired data to be used to position a blind node in a localization system composed of some landmarks. In experimental testbeds, GMM-MLE has been shown to be robust against outliers and multipath in smaller testbeds. In a testbed of size 4 m by 3 m, GMM-MLE improves the localization accuracy from 0.86 m against 1.06 m, 1.20 m and 1.58 m achieved respectively by B-MLEC, B-MLE and trilateration. However, no single algorithm outperforms all the others under all scenarios. Sometimes localization results may not be efficient on a few sole position estimation. In larger testbeds, for instance of size 8.77 m by 6.46 m, B-MLE outperforms the other localization strategies and yields to 1.94 m when three landmarks are used to compute a blind node position while 8 are considered with MLE and trilateration. Our proposed algorithms are able to improve the indoor location accuracy performance at about: 30% with GMM-MLE compared to B-MLE; 50% with GMM-MLE over trilateration localization algorithm; 40% with B-MLE over MLE and 20% with B-MLE compared to trilateration. Furthermore, such presented results also show that the best technique depends on the situation and that there is clearly a need for an online method to select the best algorithm based on a measurable parameter.
However, it is difficult to estimate such parameters in a live scenario from a mobile node's isolated point of view. Involving the architecture, relying on locations history, or even calling to machine learning techniques could be fruitful tracks that are part of our future works directions.

An immediate extension of this work is to clearly identify situations in which one algorithm is more efficient than the others and to let the nodes perform this selection on the fly. If a node is able, for instance, based on statistical properties of the sample set, to distinguish between these different situations, the resulting accuracy shall be improved. Holding a map of the network in terms of node positions can lead to know its neighboring nodes and thus a more collaboration among nodes. For example it is useful to localize itself while at the same time its neighbors. This may help to automatically determine and avoid the part of a shopping store or a parking lot garage that concentrates a lot of flow of people. Such feature also enables to intuitively determine its own position and at the same time the locations of the other conference participants in a big conference hall. In so far as sensor nodes in a wireless sensor network are mostly spread

across an unknown geographical area, the knowledge of the exact positions of its neighbors nodes is profitable for routing. For instance, the data packet can be directly forwarded to closer neighboring nodes in terms of position and thus making the routing more dynamic and more straightforward.

To this purpose, as a fourth objective, we have designed a distributed localization algorithm based on an on-line distributed stochastic approximation algorithm. This cooperative-based algorithm allows position refinement of a node which is able to estimate at the same time its own position and those of its neighbors. At each iteration of the algorithm, two uniformly randomly chosen nodes merge their common estimated positions for refinement purpose. This is different from the previous proposed algorithms in which a node is only able to estimate its own position in a sensor network. Note that, both types of scenario are possible and are applicable in real life depending on the targeted application. The RSSI-based Distributed Stochastic Approximation localization algorithm (i.e. DSAL) is compared through simulations and experiments in real environments with multilateration, multi dimensional scaling with Modern Majorization Problem (i.e. Modern MDS) and classical multi dimensional scaling. Achieved results demonstrated that the robust *DSAL* algorithm is more accurate and can improve the localization accuracy at 85% compared to MDS based localization algorithm and multilateration. A localization result of 0.77 m is obtained in a confined area of size of 5 m by 8 m at SensLAB Rennes platform using four landmarks; against 2.2 m for multilateration, 2.8 m for classical MDS and 2.4 m for modern MDS (i.e. SMACOF). Thus, the distributed DSAL localization algorithm is an attractive solution for adaptable RSSI-based localization problem under uncontrolled indoor scenario.

We have integrated and implemented statistical models and machine learning, in order to minimize the effects introduced by outliers and improve the localization accuracy. With automatic statistical filtering, we render possible the usage of such type of application: accurately localize a person at home in order to learn its habits and then to trigger several events simultaneously depending on its position. Our algorithms are initially proposed for indoor environment, but they can be applied to outdoor. In fact, the RSSI readings performed at outdoor show less variability.

6.2 Further work

Our carried research work has greatly improved the accuracy of indoor localization with the only use of the RSSI metric. We have made great efforts and the best we could to study RSSI behavior and ameliorate indoor location accuracy while ensuring incorporating diverse views including statistics and machine learning approaches. Moreover, as in any scientific research, they also open new questions and responding to them could be very valuable. In this section, we present future work and result in perspectives from the contributions of this PhD thesis.

First of all, regarding B-MLEC localization algorithm (cf. Chapter 4), the choice of the threshold of acceptance-rejection of a landmark was determined experimentally by comparing the results obtained by different rejection threshold values. We have experimentally determined the order of the threshold that gives the strongest results in term of location accuracy, i.e. the smallest distance error. An immediate extension is to provide a theoretical study in order to bind the step of acceptance-rejection of a landmark to the general methodology of statistical test. This is to assess whether the implemented decision policy follows in fact, a particular form of statistical test (e.g. famous Neyman-Pearson lemma). If this is the case, it would give us the theoretical method for choosing the rejection threshold associated with the optimal decision

policy of acceptance-rejection of a landmark.

Another extension of this work is multi-hop networks, in which such an approach can help to form groups of coherent nodes that can reinforce their mutual location. Different groups can then exchange and combine information e.g. the propagation model parameters to enhance global positioning. Such technique allows processing data locally as it is closer to the source. Also, since sensors will communicate data associated with a small distance in a group/cluster environment, the energy consumed by the network will be lower than the one consumed without group/cluster formation in the network.

In addition, the network self organization properties of sensor networks can be exploited through the grouping of node or "clustering" to provide an efficient mechanism to manage addressing and Internet Protocol version 6 (i.e. IPv6) addresses prefixes. Sensor nodes can thus be grouped according to certain criteria. A group formation can be detailed as follows: a group leader also called cluster head can be selected based on its location or proximity in relation to a number of close neighbors, its identifier or its residual energy. The IP prefix can then be associated with the position of the cluster head to form a stable information shared by nearby nodes or nodes which move together. Therefore a cluster head can handle the distribution of addresses of its group according to an assigned prefix.

Rendering every sensor node sensitive to its position (absolute or relative), allows the design of an automatic and fully distributed addressing mechanism, facilitating the deployment of sensor networks e.g. for routing protocols. That means no manual address configuration; neither a central server for address management. Indeed, tracking nearby sensor nodes allows the formation of close sensor nodes clusters and thus, enable to route data cluster among cluster.

The routing can be thus identified macroscopically, by sending the information to the group (cluster head node or another node in the group). Then any node in the group receiving the data can route them to the next group. This allows scalability and also contributes to limit sending routing packet in the entire network.

To belong to a cluster, a node having a knowledge of its position, chooses the closest cluster head in terms of location. Moreover, once established, the grouping of nodes can be further exploited using the results of data aggregation to locate a comer node with unknown position. Thus, a node can easily and dynamically acquire an address by means of membership in a group, i.e. its position. If the node happens to move, this address will be freed and can be reused by another node. In the case where a node moves, this last should acquire a new IP address as its new position will be probably related to a new group/cluster. An application such as a routing protocol uses these IP addresses for data communication. A blind mobile node location/path can be exploited by a routing protocol for authentication purpose before sending data. The routing protocol thus cooperates actively with the localization and clustering algorithms. The clustering algorithm should then dynamically manage and in real-time the mapping between IP address $\leftrightarrow$ node identifier. Such information is communicated to the routing protocol as presented in Figure 6.1.

Moreover, the deployment of these sensor nodes capable of localizing themselves in a smart home application will foster an environment that is open to heterogeneous equipment including smart phones, computers, home automation actuators: a set of objects that communicate and collaborate. This contributes to have a distributed localization architecture which can be opportunistic and heterogeneous. In addition of accepting the node mobility, this localization architecture must be able to locate a node based on available information. Thus, such localization architecture should be able to choose the best localization algorithm according to the propagation

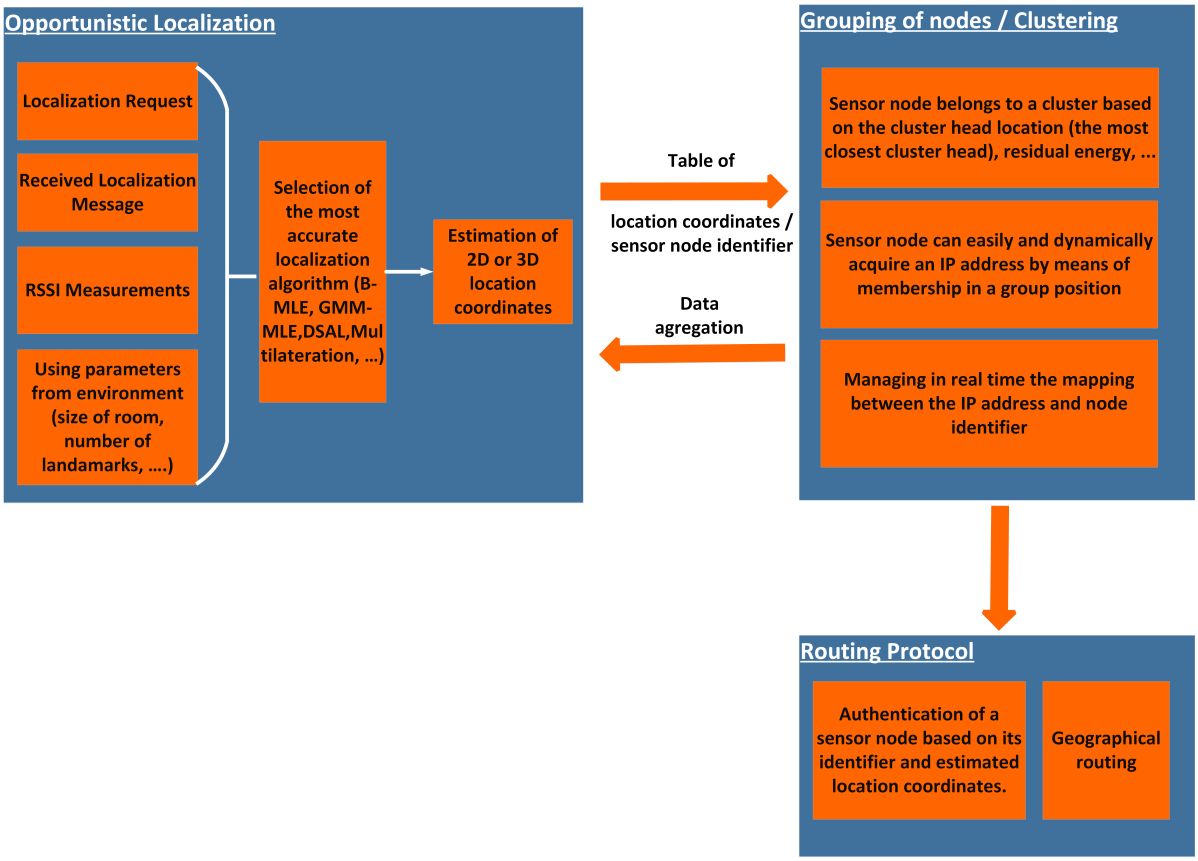


Figure 6.1: Proposed architecture for clustering-based location to enhance routing in multihop wireless sensor network

conditions of the environment (room size, number of landmarks, size of the *learning set* for the computation of the propagation model parameters, ranging errors, etc.). Furthermore, this opportunistic localization system can communicate with home automation actuators to trigger actions depending on the position of the occupant. Such system can additionally collaborate with smartphone for event notification to the occupant.

The energy consumption constraints of sensor nodes should also be taken into account in order to maximize the lifetime of the sensors network. This can be done by programming the sensors in inactive mode when no localization request is received by adding features to detect the presence of a user. The size of the network should also be considered due to the fact that, more the network is bigger and more there are messages. At long-term, better living quality and location-based services can be supplied to smart home users such as services related to home automation by automatically selecting the appropriate localization algorithm that is the best depending on the environment characteristics and required services.

Finally, it must be mentioned that the security was not a major concern when designing the proposed algorithms. However, the communicating micro-components that use these algorithms function in wireless mode and are targeted to multiple attacks. This requires special attention to detect an intrusion that may happen on the communication channel as this last is not currently encrypted and thus data packets can be modified. Hackers can also intrude the communicating node itself by recovering its contents. Along these lines, it is important to add security aspects on each localization algorithm. For example, this can be done by exploiting the mobility, because a trajectory is a unique feature for each person. Localization algorithms should also include dynamic cryptography key that are also energy efficient as besides the localization, the energy consumption is one of the major problems for sensor networks.

Synthèse en français

Introduction

L'internet des objets se développe à travers diverses applications comme la domotique, la surveillance à domicile, etc.. Les consommateurs s'intéressent à ces applications dont les objets interagissent avec des dispositifs de plus en plus petits et connectés.

La localisation est une information clé pour plusieurs services ainsi que pour l'optimisation du fonctionnement du réseau. En environnement intérieur ou confiné, elle a fait l'objet de nombreuses études. Cependant, l'obtention d'une bonne précision de localisation demeure une question difficile, non résolue. Tout algorithme de localisation utilise des nœuds avec positions connues appelés landmarks et des nœuds dont les positions sont à localiser. Pour le GPS, les landmarks sont les satellites. Mais comme nous le savons tous, les signaux GPS sont fortement atténués en environnement interne. En effet, même si en ville nous avons des immeubles, l'environnement de propagation interne est beaucoup plus hostile. Le signal sans fil en environnement interne est fortement atténué et a un comportement imprévisible du fait de la présence d'obstacles en intérieur du bâtiment mais également la façon dont les objets sont disposés dans l'environnement.

Cette thèse étudie le problème de la localisation en environnement intérieur appliqué aux réseaux sans fil avec l'utilisation unique de l'atténuation du signal. L'atténuation est mesurée par l'indicateur de l'intensité du signal reçu (RSSI). Le RSSI est connu dans la littérature comme étant imprécis et peu fiable en ce qui concerne l'estimation de la distance, du fait de la complexité de la propagation radio en intérieur : il s'agit des multiples trajets, le shadowing, la fading. Cependant, il est la seule métrique directement mesurable par les petits objets communicants et intelligents. La mesure du RSSI ne nécessite ni la synchronisation d'horloge entre un émetteur et un récepteur radio comme pour la mesure du temps d'arrivée du signal (ToA), ni la mise en œuvre de matériel supplémentaire comme le requiert la mesure de l'angle d'arrivée du signal (AOA), ce qui est prohibitif et consomme beaucoup d'énergies.

Pour nos travaux, nous nous sommes intéressés à l'amélioration de la précision des mesures du RSSI pour les rendre applicables à l'environnement interne pour une localisation plus précis. Nous nous sommes également intéressés à l'implémentation et au déploiement de solutions algorithmiques relatifs au problème suivant : est-il possible d'obtenir une meilleure précision de localisation en utilisant uniquement la mesure du RSSI fournie par les nœuds capteurs sans fil IEEE 802.15.4 ?

La plupart des travaux de l'état de l'art se focalisent à proposer des algorithmes coopératifs dans le but d'améliorer la précision de localisation sans se préoccuper du modèle sous-jacent de la propagation radio, un facteur qui affecte l'erreur de localisation. Dans cette perspective, nous avons d'abord étudié le comportement du RSSI dans plusieurs environnements intérieurs de différentes tailles et selon plusieurs configurations, y compris un réseau de capteurs sans fil à grande échelle (SensLAB). Pour expliquer les résultats des mesures, nous avons caractérisé les objets communicants que nous utilisons dans nos expériences, les nœuds capteurs Moteiv TMote Sky par une série de mesures en chambre anéchoïque. Nos études expérimentales confirment que la relation entre le RSSI et la distance dépend de nombreux facteurs même si la batterie intégrée à chaque nœud capteur produit une atténuation. Une fois le nœud capteur sans fil qualifié, nous avons mieux compris le comportement du RSSI.

Ensuite, nous avons démontré que le modèle de propagation log-normal shadowing n'est pas adapté en intérieur, en raison de la dispersion des valeurs de RSSI et du fait que celles-ci ne sont pas toujours dépendantes de la distance. Il en résulte que ces valeurs devraient être considérées séparément en fonction de l'emplacement de chaque nœud capteur émetteur sans fil. Ainsi, les résultats expérimentaux ont mis en évidence que la technique du RSSI n'aboutit pas toujours à une bonne précision d'estimation de la distance dans un environnement intérieur réaliste.

Nous avons proposé des heuristiques pour corriger ces incohérences observées. Nos algorithmes tiennent compte des effets de la propagation par trajets multiples et des valeurs aberrantes. Les effets de la propagation par trajets multiples sont pris en compte en biaisant le modèle classique de propagation log-normal shadowing. Biaiser le modèle de propagation log-normal shadowing conduit à sélectionner les landmarks qui ont un comportement anormal. Avec cet algorithme de localisation nommé B-MLE, un landmark est considéré comme aberrant ou anormal lorsque les mesures de RSSI associées n'améliorent pas la précision de la localisation estimée par le nœud mobile aveugle. L'algorithme de localisation B-MLE est aussi comparé à une autre proposition qui sélectionne efficacement les landmarks normaux sur la base d'un seuil. La valeur du seuil est choisie de façon empirique sur la base d'expérimentations.

Nous avons aussi développé un nouvel algorithme de localisation s'appuyant sur le RSSI et utilisant une approche statistique d'apprentissage automatique de données dont le modèle de mélanges Gaussiens à deux modes. Cet algorithme de localisation nommé GMM-MLE est un procédé à deux niveaux : il identifie d'abord la distribution des échantillons de données de mesures de RSSI comme la composition d'un grand pourcentage de valeurs dites « normales »et d'une faible proportion de valeurs aberrantes. Ensuite la position d'un nœud mobile aveugle est localisée en appliquant l'estimateur du maximum de vraisemblance conjointement avec l'approche statistique de modèle de mélanges Gaussiens à deux modes.

Les algorithmes de localisations proposés pour prendre en compte les incohérences observées du milieu interne à savoir la propagation par trajets multiples et la présence de valeurs aberrantes (les méthodes basées) sont comparés aux travaux de l'état de l'art : la trilatération classique sur l'ensemble des données et le maximum de vraisemblance appliqué au modèle log-normal shadowing appelé encore MLE non biaisé. Nos résultats expérimentaux obtenus en environnements internes réels ont confirmé que nos algorithmes améliorent significativement la précision de localisation en intérieur avec utilisation de la seule métrique du RSSI.

Enfin, nous avons étudié et proposé un algorithme de localisation distribué précis qui passe à l'échelle, respecte les caractéristiques de l'informatique embarquée des réseaux de capteurs sans fil et peu consommateur en termes de temps de calcul. L'algorithme de localisation d'approximation stochastique distribué (DSAL) proposé affine la précision des positions estimées par un nœud capteur sans fil en utilisant la technique du RSSI. Notre approche a été comparée à d'autres algorithmes distribués de l'état de l'art. Les résultats issus des simulations et des expériences réalisées dans des environnements internes réels ont révélé une meilleure précision de la localisation avec notre algorithme distribué. L'erreur de localisation obtenue est de l'ordre du centimètre sans aucun nœud ou unité centrale de traitement ni de calibration fastidieuse ni d'intervention humaine.

Contributions

Dans la première partie de cette thèse nous allons présenter les résultats de campagnes de mesures pour caractériser le comportement du RSSI en interne. Il s'agit du diagramme de rayonnement du

nœud capteur Tmote Sky que nous pensons représentatif. Nous étudions ensuite le comportement du RSSI en fonction de la distance en environnement réel après avoir validé l'utilisation du modèle log-normal shadowing en environnements internes.

Dans la deuxième partie de cette thèse, nous allons ensuite présenter les mécanismes et les résultats obtenus par les algorithmes de localisation à un saut que nous avons proposé. Notre démarche se base sur des simulations et des expériences.

Dans la dernière partie de cette thèse, nous allons présenter l'algorithme stochastique distribué que nous avons proposé pour affiner l'ensemble des positions des nœuds dans un réseau de capteur sans fil. Nous avons évalué les performances de notre algorithme dans la plateforme SensLAB de Rennes. Les résultats issus des simulations et des expériences en environnements réels obtenus par l'algorithme DSAL ont été comparés aux résultats obtenus par les algorithmes distribués de l'état de l'art : la multilatération, le multi dimensional scaling with modern majorization problem et le classical multi dimensional scaling. Ces résultats ont révélé une meilleure précision de la localisation basée sur l'approximation stochastique distribuée (DSAL). Nous allons finir par conclure sur les travaux futurs.

Dans les sections qui suivent nous présentons brièvement chacune des parties mentionnées de la thèse.

Caractéristiques du modèle de propagation

Nos travaux se basent sur les algorithmes à mesure de distance ou « range-based algorithm », plus précisément ceux qui utilisent un modèle de propagation radio. En d'autres termes, il sera nécessaire de construire l'équation d'un modèle de régression sur la relation entre les valeurs de RSSI mesurées et la distance correspondante. Ceci afin d'estimer les distances entre chaque nœud landmark et la source aveugle à localiser.

Le modèle de propagation le plus couramment utilisé en environnement interne et externe et qui représente un bon compromis entre faisabilité et complexité est le modèle log-normale shadowing (LNSM) proposé par Rappaport et Seidel [22]. C'est un modèle empirique qui prend en compte les effets de réverbération conséquence des perturbations affectant le signal le long du canal de communication entre un émetteur et un récepteur radio. La réverbération ou « shadowing »est définie par le fait qu'en milieu interne, le signal ne se propage pas directement entre un émetteur et un récepteur, mais se diffuse sur toute la surface de l'environnement interne. Par environnement interne, on sous-entend une pièce fermée, des endroits confinés, sous un tunnel.

L'utilisation du modèle log-normale shadowing requiert une distribution log-normale (comme présenté sur la figure 6.2) de la puissance reçue à n'importe quelle distance de l'émetteur. La fonction mathématique du modèle log-normale shadowing est la suivante (cf. equation (6.1)) :

$$\mathrm{PL}(d) = \mathrm{PL}(d_0) + 10.\eta.\log_{10}\left(\frac{d}{d_0}\right) + \mathcal{N}(0,\sigma^2)\,, \tag{6.1}$$

Ce modèle de propagation log-normal shadowing exprime l'atténuation $\mathrm{PL}(d)$ entre un émetteur et un nœud récepteur radio séparé d'une distance d que nous voulons estimer. Avec $\mathrm{PL}(d_0)$ la puissance de référence exprimée en « dBm »et mesurée à la distance de référence d_0 (généralement égal à 1 m en environnement interne. η est le facteur prenant en compte l'atténuation de l'environnement. $\mathcal{N}(0,\sigma^2)$ est une variable aléatoire gaussienne de valeur moyenne nulle et

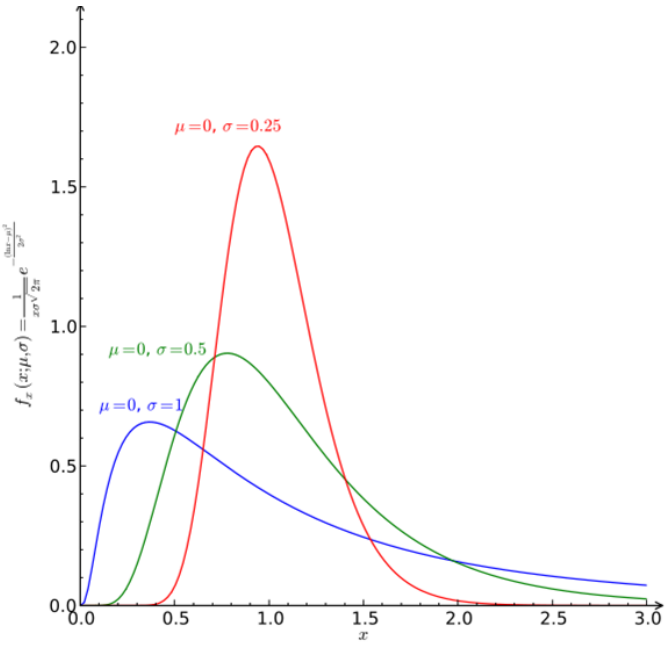


FIGURE 6.2 – An example of log-normal distribution [201] [202]

d'écart-type σ. η est le paramètre de perte de trajet et σ^2 prend en compte les effets de réverbération (en anglais « shadowing ») dans l'environnement.

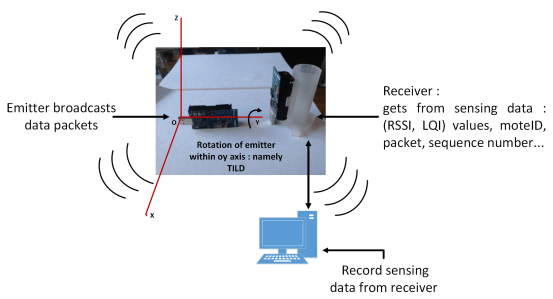
Ce modèle ne tient pas compte des effets de « fast fading »ou variations brusques et rapides dans la mesure où plusieurs valeurs de RSSI peuvent être lues à partir des messages provenant d'un même nœud émetteur radio [5]. Dans la plupart de nos expériences, nous utilisons une puissance d'émission de $0\,dBm$, le RSSI est de ce fait égal à l'opposé de l'atténuation mesurée : $RSSI(d) = -\mathrm{PL}(d)$.

Etude du comportement du RSSI en environnements internes réels

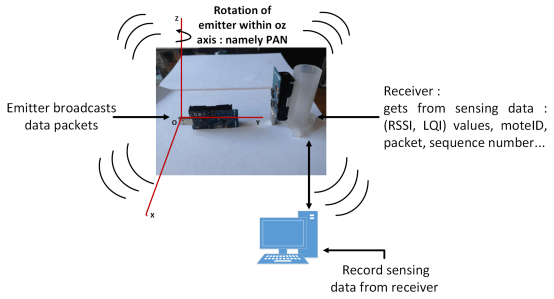
Cette section caractérise les nœuds capteurs Tmote Sky en chambre anéchoïque mais également en environnements réels. Une chambre anéchoïque est dépourvue de réflexions, d'interférences car composé de matériaux absorbant les ondes électromagnétiques. Notre objectif est de caractériser le gain de l'antenne dans la chambre anéchoïque mais également en environnements internes réels avec des perturbations supplémentaires dans le but de mieux caractériser le comportement du RSSI.

Diagramme de rayonnement du nœud capteur Tmote Sky dans une chambre anéchoïque

Pour déterminer les positions relatives du nœud capteur Tmote Sky les plus favorables ou défavorables au bilan de la liaison, nous avons réalisé son diagramme de rayonnement dans une chambre anéchoïque. Nous avons mesuré les valeurs de RSSI sur un récepteur en position de déploiement horizontal depuis les messages envoyés par un émetteur en position de déploiement horizontal. Ainsi nous faisons d'abord tourner l'émetteur suivant les axes horizontaux puis

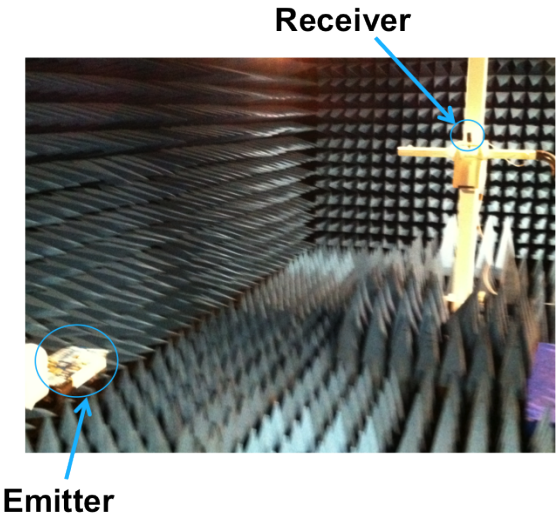


(a) Rotation de l'antenne d'un nœud capteur émetteur suivant l'axe horizontal

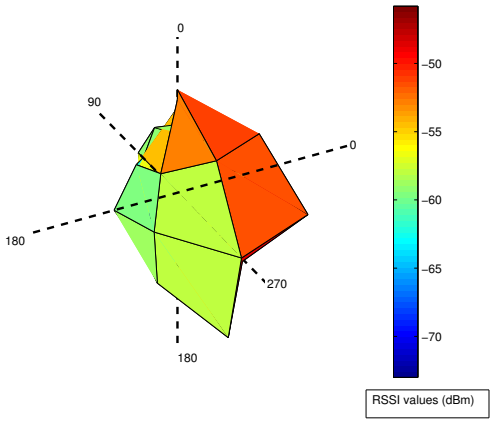


(b) Rotation de l'antenne d'un nœud capteur émetteur suivant l'axe vertical

FIGURE 6.3 – Architecture du diagramme de rayonnement 3D du nœud capteur Tmote Sky réalisé dans une chambre anéchoïque



(a) Une vue de dessus de la chambre anéchoïque



(b) Diagramme de rayonnement 3D du nœud capteur Tmote Sky réalisé dans une chambre anéchoïque

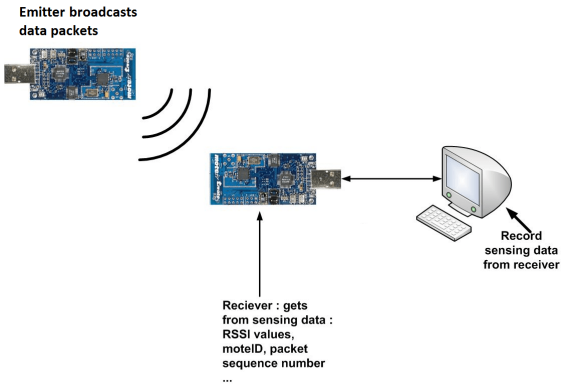
FIGURE 6.4 – Diagramme de rayonnement 3D du nœud capteur Tmote Sky réalisé dans une chambre anéchoïque

suivant les verticaux pour un déploiement vertical du récepteur.

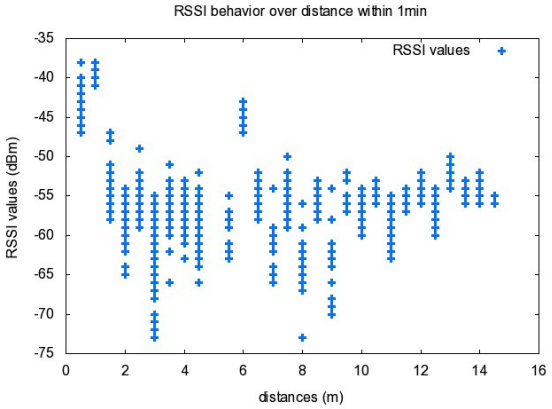
Les résultats ont démontré un diagramme proche mais qui n'est pas tout à fait omnidirectionnel. En effet nous recherchons une grande plage de valeurs où la directivité est assez constante pour une plage angulaire la plus grande possible, ce qui n'est pas le cas dans le graphique ci-dessous sur lequel nous avons remarqué des disparités.

Le diagramme de rayonnement du nœud capteur Tmote Sky réalisé dans la chambre anéchoïque de Telecom ParisTech ne montre ni une forme directive ni un modèle totalement omnidirectionnelle : il ne rayonne pas de la même manière sur toutes les directions.

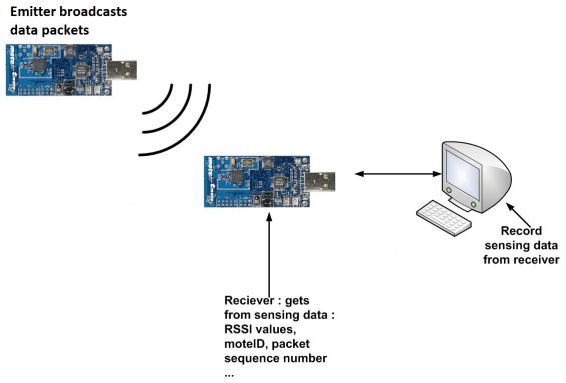
Le graphique de droite présente le diagramme de rayonnement obtenu. La partie de droite du diagramme présente une meilleure polarisation lorsque l'angle horizontal est d'environ 0°, -90° et 270° pour tout émetteur respectant la position de déploiement comme montré sur le graphique de gauche. Plus de −50 dBm comme valeurs de RSSI sont observées sur cette plage angulaire, ce qui



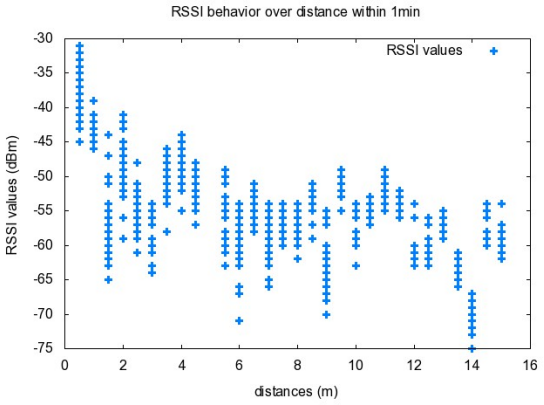
(a) Scénario de l'architecture TowardCenter parallel antennas (i.e A)



(b) Les valeurs de RSSI mesurées du scenario A



(c) Scénario de l'architecture FromCenter parallel antennas (i.e B)



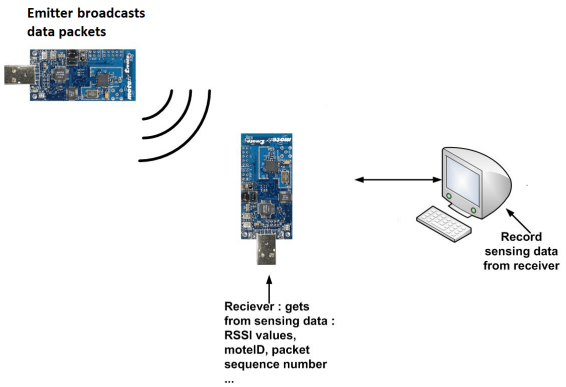
(d) Les valeurs de RSSI mesurées du scenario B

FIGURE 6.5 – Distribution des valeurs de RSSI mesurés pour différentes positions de déploiement de capteurs émetteurs-récepteur sur le testbed Lincs (partie 1)

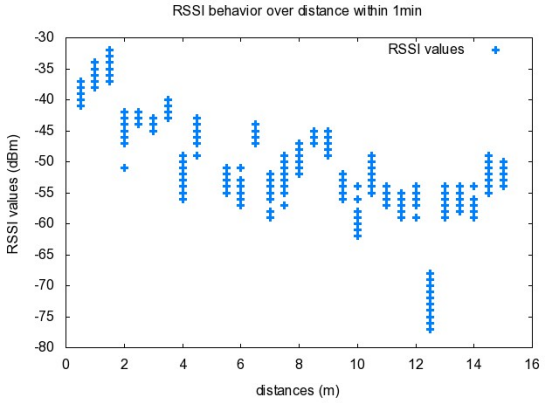
est mieux comme polarisation qu'aux alentours de 180° vers lesquels on observe des valeurs de RSSI comprises entre −60 dBm et −55 dBm. Par ailleurs, quand on regarde les angles verticaux, nous remarquons des valeurs de RSSI comprises de -50 dBm et −60 dBm à −55 dBm à 180°. Nous remarquons que l'orientation de l'émetteur par rapport à la position de déploiement du récepteur est importante.

Nous avons réalisé des tests sur différentes orientations possibles et par rapport à notre démarche, le fait de changer d'orientation n'influence pas significativement les résultats, même si cette position de déploiement dénommée *TowardCenter perpendicular* obtient un comportement plus proche du comportement log-normal shadowing comparé aux autres (i.e. *TowardCenter parallel*, *FromCenter perpendicular* et *FromCenter parallel*) présenté et interprété dans la sous section 3.4.2 du chapitre 3.

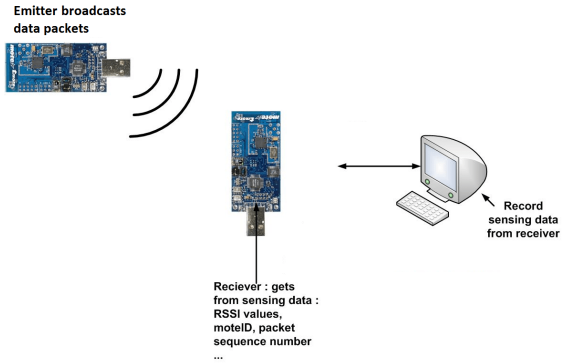
Les résultats du diagramme de rayonnement 3D du nœud capteur Tmote Sky sont présentés pour une position de déploiement vertical d'un nœud capteur récepteur par rapport à un déploiement horizontal de nœuds capteurs émetteurs. Le diagramme de rayonnement 3D obtenu n'est pas tout à fait omnidirectionnel, mais nous pouvons l'utiliser tout en sachant que cette limite existe.



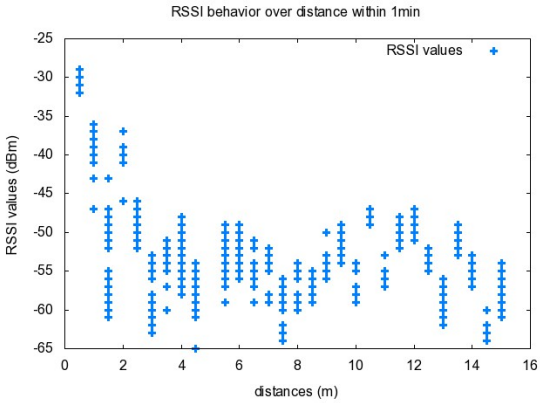
(a) Scénario de l'architecture TowardCenter perpendicular antennas (i.e C)



(b) Les valeurs de RSSI mesurées du scenario C



(c) Scénario de l'architecture FromCenter perpendicular antennas (i.e D)



(d) Les valeurs de RSSI mesurées du scenario D

FIGURE 6.6 – Distribution des valeurs de RSSI mesurés pour différentes positions de déploiement de capteurs émetteurs-récepteur sur le testbed Lincs (partie 2)

Dans la suite nous allons étudier le comportement du RSSI en environnements réels avec des perturbations.

Comportement du RSSI en environnements internes réels avec des perturbations

Après avoir caractérisé les nœuds capteurs, nous étudions le comportement du RSSI en environnements réels sous la présence d'obstacles.

Nous avons mesuré les valeurs de RSSI entre un émetteur et un récepteur radio qui s'éloigne dans un testbed, un bureau meublé et occupé de taille 7 m par 5 m situé dans le laboratoire Lincs, département Infres, Télécom ParisTech Place d'Italie. Nous avons remarqué que pour une même position de déploiement entre l'émetteur et le récepteur, dépendant de la manière dont l'émetteur et le récepteur sont orientés en respectant ou non la proximité avec le mobilier, nous n'obtenons pas le même comportement des valeurs mesurés de RSSI par rapport à la même distance.

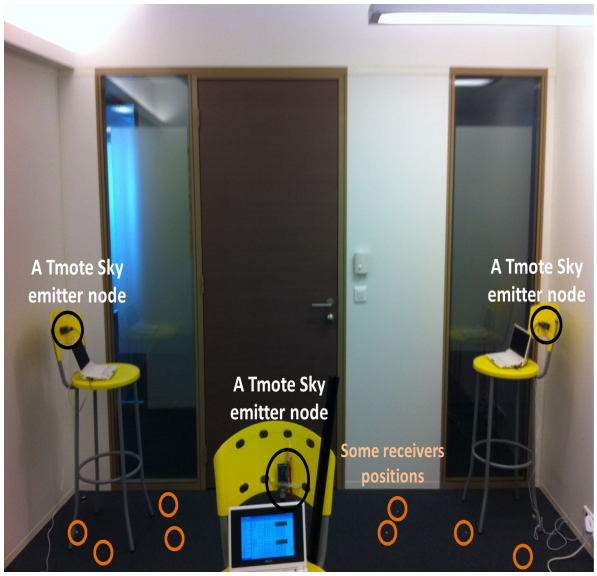
Ces deux graphiques représentent la distribution des valeurs de RSSI en fonction de la distance. Nous observons un comportement proche de la distribution log-normale sur le graphique de gauche où nous avons mesurées les valeurs de RSSI entre un émetteur se trouvant près du mobilier et un récepteur qui se déplace du mobilier vers le fond de la pièce (i.e. le mur).

Cependant, nous n'observons pas la même dispersion des valeurs de RSSI mesurées pour les mêmes distances lorsque l'émetteur se trouve au fond de la salle (i.e. le mur) et que le récepteur se déplace vers le mobilier (cf. graphique de droite). Ces résultats démontrent que le mobilier a un effet perturbateur sur le signal radio sans fil en interne. En effet, sans perturbation, nous aurions dû obtenir le même comportement au niveau de la dispersion des valeurs de RSSI puisque les mesures de RSSI sont réalisées aux mêmes distances et suivant une même position de déploiement entre un émetteur et un récepteur.

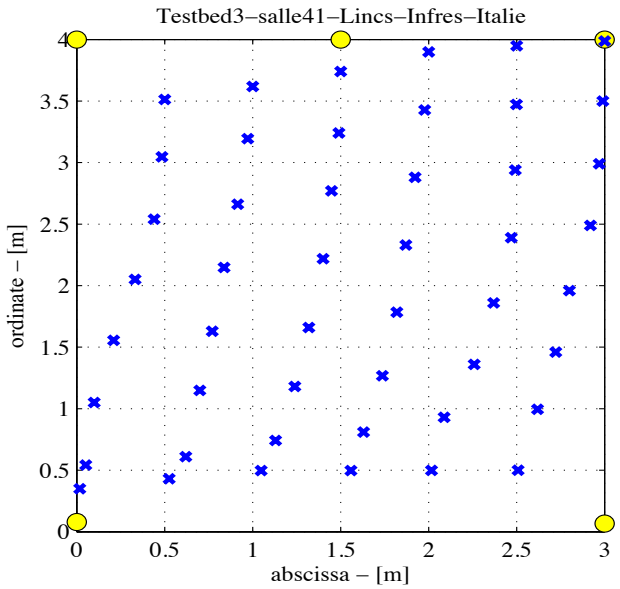
Les résultats obtenus suggèrent la nécessité de prendre en compte l'orientation des antennes des émetteurs (ou encore landmarks) par rapport à celle du nœud aveugle à localiser. Ce qui est fait par nos algorithmes même si nous ne prenons pas cela en compte explicitement sur les nœuds. En effet, nous n'imposons pas une orientation particulière aux antennes émettrices et réceptrices. Il faut noter que dans nos travaux nous avons utilisé une seule position de déploiement du récepteur sur lequel nous mesurons les valeurs de RSSI depuis les messages envoyés par tout émetteur en position de déploiement horizontal. Toutefois, nous avons également Ã©tudié d'autres orientations (cf. chapitre 3, section 3.4.2) qui montrent des formes différentes des courbes des valeurs de RSSI en fonction de la distance mais décrivant toujours les mêmes tendances même si l'antenne du nœud capteur Tmote Sky n'est pas isotrope.

Présentation des différents testbeds

Cette section présente une vue de dessus de nos différents testbeds. Dans les différents graphiques des figures 6.7 à 6.9 modélisant les tailles des testbeds, les landmarks sont illustrés par les cercles jaunes et les positions du nœud aveugle à localiser par les croix bleues. Le graphique de gauche présente le plus petit dénommé Lincs et différent du testbed précédemment mentionné. Le testbed lincs est un bureau de taille 4 m par 3 m, situé au laboratoire Lincs, Telecom Paristech



(a) Une vue de dessus du testbed Lincs (#1) de taille 4 m * 3 m)



(b) Testbed Lincs : (5 landmarks, 48 positions)

FIGURE 6.7 – Position des landmarks (les cercles jaunes) et de la source à localiser (les croix bleues) sur le testbed Lincs (#1)

site place d'Italie. Le testbed Lincs se compose de 48 positions du nœud mobile aveugle à localiser et de 5 landmarks représentés par les cercles jaunes. Le graphique de droite quand à lui représente le testbed nommé Barrault, qui est la salle C49 barrault. Le testbed Barrault mesure $7m \times 7m$: il se compose de 47 positions du nœud mobile aveugle à localiser et de 4 landmarks. Nous avons également le testbed Rammus, qui est une salle de classe (A27) se trouvant sur la plateforme Rammus, Télécom Bretagne Rennes. Il est représenté par le graphique de droite de la figure 6.9. Il se compose de 8 landmarks et de 57 positions de la source à localiser. Dans chaque testbed nous avons en moyenne la présence de deux personnes et des points d'accès sans fil. Chacune des positions de test (représentées par les croix bleues) est choisie au hasard et différente pour chaque testbed. Ainsi, sur chacune de ces positions de test, la source à localiser est déplacée par un opérateur et estime sa position selon les valeurs de RSSI mesurées sur les paquets envoyés par chaque landmark selon la figure 6.14. La localisation est décidée instantanément et permet ainsi la mobilité.

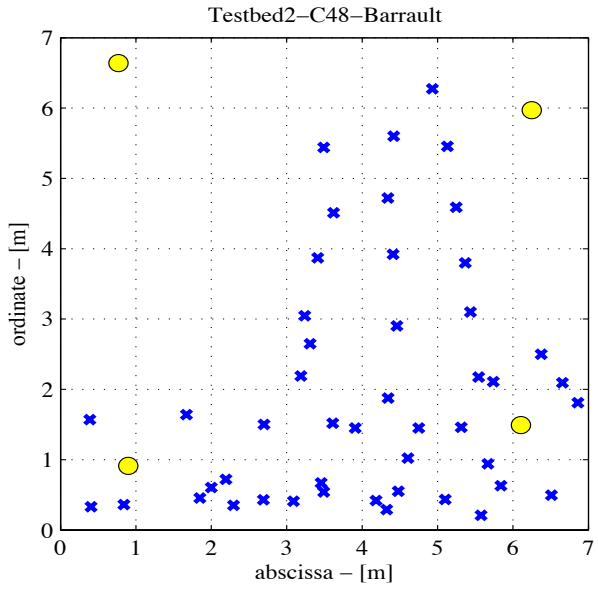
Enfin, nous avons le testbed SensLAB (cf. figure 6.9) de taille 5 m par 8 m, une plateforme de capteurs sans fil à large échelle se trouvant sur la plateforme SensLAB du site de Rennes. Le testbed SensLAB utilise 12 nœuds capteurs aveugles avec des positions inconnues et 4 landmarks.

Influence des trajets multiples sur le RSSI

Dans les sections ci-dessus nous avons caractérisé le nœud capteur Tmote Sky dans une chambre anéchoïque et en environnements internes réels. Dans ce paragraphe nous allons étudier la propagation du signal radio sans fil en environnement réel.

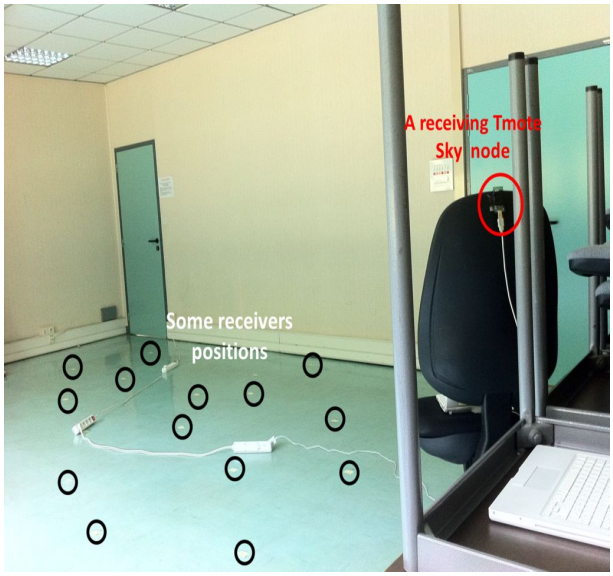
Nous avons représenté sous forme d'histogramme la distribution des valeurs de RSSI mesurées

(a) Une vue de dessus du testbed Barrault (#2) de taille 7 m par 7 m)

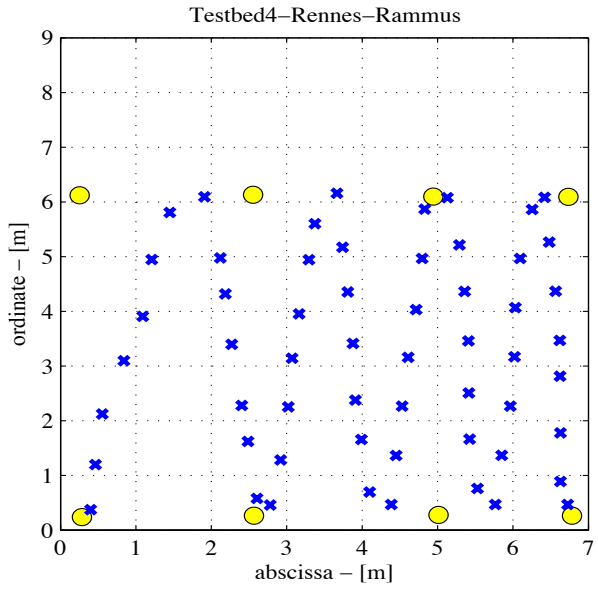


(b) Testbed Barrault : (4 landmarks, 47 positions)

FIGURE 6.8 – Position des landmarks (les cercles jaunes) et de la source à localiser (les croix bleues) sur le testbed Barrault (#2)

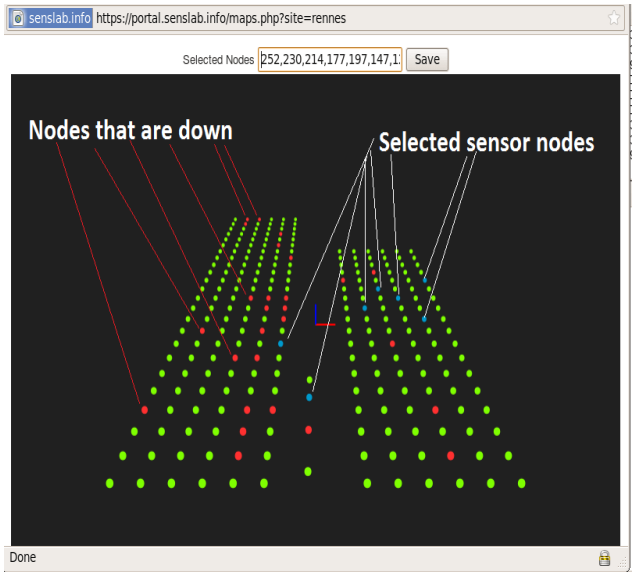


(a) Une vue de dessus du testbed Rammus (#3) de taille 8.77 m par 6.46 m)

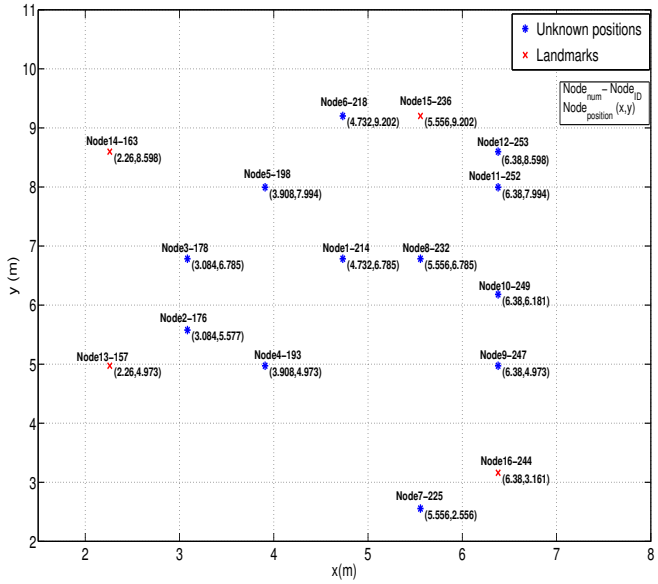


(b) Testbed Rammus : (8 landmarks, 57 positions)

FIGURE 6.9 – Position des landmarks (les cercles jaunes) et de la source à localiser (les croix bleues) sur le testbed Rammus (#3)



(a) Une vue de dessus du testbed SensLAB Rennes. Un réseaux de capteurs sans fil



(b) Testbed SensLAB Rennes : (4 landmarks, 12 positions de nœuds aveugles)

FIGURE 6.10 – Testbed SensLAB Rennes : position des landmarks représentés par les croix rouges et des nœuds aveugles à localiser symbolisées par les astérisque bleues

sur une période du temps, sur une position du récepteur en fonction de l'emplacement de chacun des 8 landmarks (depuis les messages envoyés) dans le testbed Rammus. Nous avons représenté le nombre de valeurs de RSSI mesurés sur les paquets de données reçus avec succès de chaque landmark. La figure 6.11 présente les valeurs de RSSI en abscisse et le pourcentage des valeurs de RSSI en ordonnée. Nous remarquons que les distributions des valeurs de RSSI pour chaque landmark en ce qui concerne le récepteur ne sont pas les mêmes ; ceci est normal car les distances entre chaque landmark et cette position arbitraire du récepteur ne sont pas les mêmes.

Toutefois, nous devrions mesurer de petites ou égales variations sur l'écart type parce que l'environnement ne change pas et les conditions de propagation sont les mêmes. D'où l'idée de considérer que les paramètres du modèle de propagation log-normal shadowing (i.e. LNSM) dépendent de l'emplacement des landmarks.

Cet histogramme de la figure 6.11 montre que le facteur dominant est la propagation par trajets multiples. En fait, notre étude montre clairement que le nœud émetteur dont les valeurs (de RSSI) mesurées sur le récepteur et représentées à la deuxième position à partir du haut est un peu loin en termes de dispersion des valeurs RSSI à l'égard des autres émetteurs. En effet, ce nœud émetteur dont les valeurs de RSSI sont situées à la deuxième position à partir du haut sont trop dispersées par rapport à l'ensemble des autres valeurs mesurées des autres landmarks. A cause des trajets multiples, cet émetteur se trouve en mauvaises conditions de propagation par rapport à la position du récepteur, entrainant de larges dispersions qui sont observées au niveau des valeurs de RSSI mesurées sur le récepteur des messages envoyés par cet émetteur. Dans le chapitre 3, nous avons aussi démontré que la sensibilité au niveau du récepteur existe, mais elle ne peut pas causer d'aussi larges variations. De plus, les trajets multiples impliquent également que les valeurs de RSSI mesurés à partir des messages envoyés par un landmark sont soient trop fortes, soient trop faibles même à courte et égale distance (cf. figure 6.12).

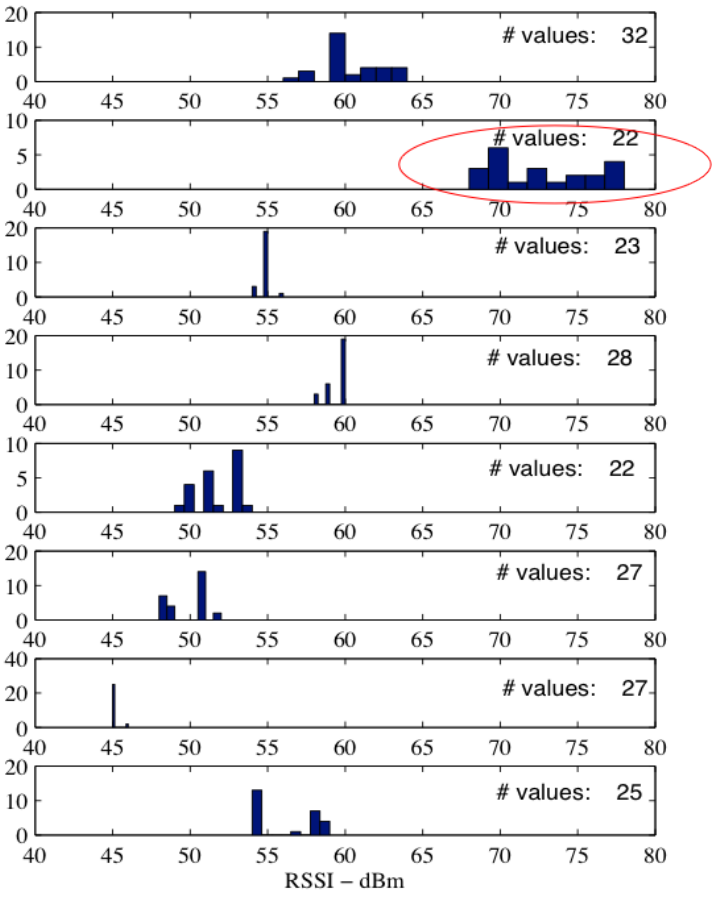


FIGURE 6.11 – Histogramme des valeurs de RSSI mesurées sur un émetteur à partir des messages provenant de 8 landmarks différents sur le testbed Rammus

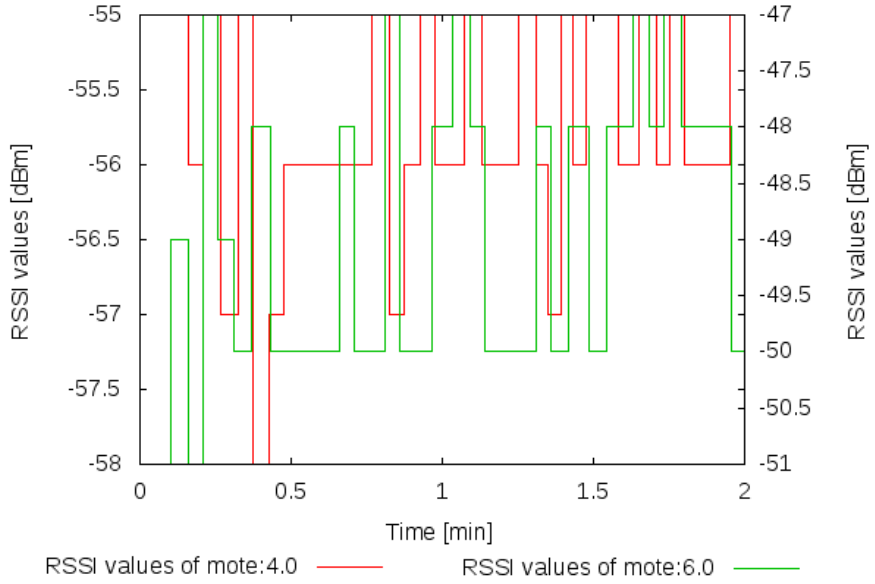


FIGURE 6.12 – Deux comportements différents de RSSI mesurés sur deux nœuds distincts situés à égale distance de 5.6 m d'un nœud émetteur dans un bureau semi-meublés

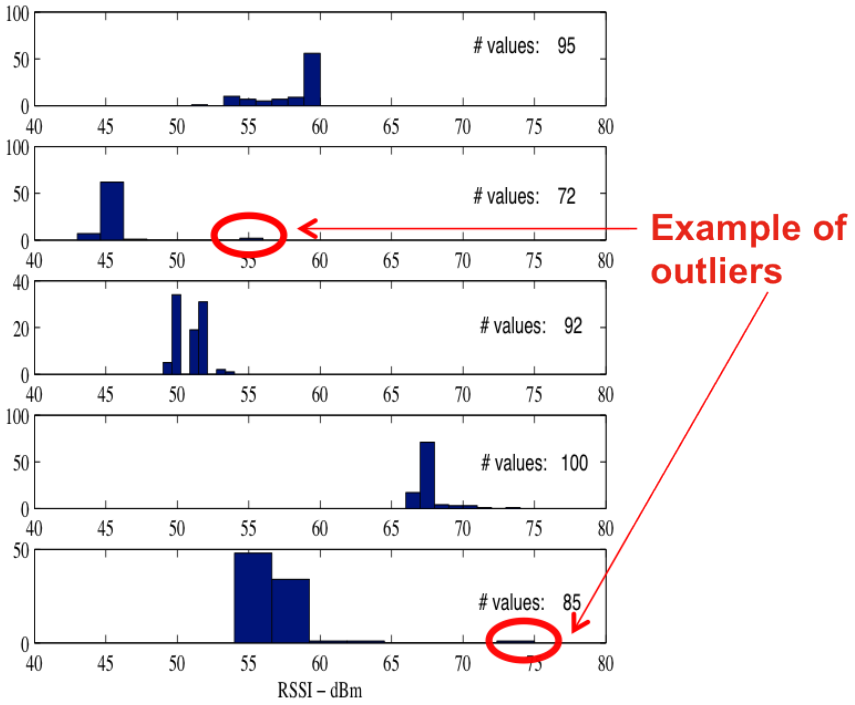


FIGURE 6.13 – Histogramme des valeurs de RSSI mesurés sur un émetteur à partir des messages provenant de 5 landmarks différents sur le testbed Lincs

Nous pouvons donc considérer qu'un nœud landmark parmi ceux du réseau peut être affecté d'un comportement anormal. De ce fait les valeurs de RSSI mesurées sur un récepteur à partir des messages envoyés par ce landmark sont considérées comme étant corrompues.

Nous avons également illustré la distribution des valeurs de RSSI mesurées sur une position du récepteur en fonction de l'emplacement de chacun des 5 landmarks sur le testbed Lincs (cf. figure 6.13). Nous remarquons cette fois la présence de valeurs aberrantes. Ceci est peut-être normal du fait de la petite taille du testbed qui le rend favorable ou plus sensible aux interférences.

Les résultats de ces expériences et d'autres que nous avons réalisés au laboratoire Lincs et au départment Infres Telecom ParisTech montrent clairement que les facteurs dominants sont les trajets multiples (en anglais « multipath ») et les valeurs aberrantes (« outliers »).

Description du processus de localisation

La figure 6.14 présente notre processus de localisation tel qu'il est réalisé. Sur chaque position de test, le nœud mobile calcule instantanément sa position en utilisant les valeurs de RSSI reçues des messages envoyés par chaque landmark et la position de ces derniers.

Estimation en ligne des paramètres du modèle de propagation

D'un point de vue de notre architecture de localisation, tout en recevant des données nécessaires au calcul de sa position, le nœud aveugle calcule instantanément les paramètres du modèle de

FIGURE 6.14 – Une vue de notre processus de localisation réalisé instantanément sur chaque position

propagation relative à son environnement. Pour réaliser cela, le nœud se base sur les équations du modèle de régression basées sur le maximum de vraisemblance. Ce qui est différent des autres travaux de l'état de l'art pour lesquels les estimations des paramètres du modèle de propagation sont réalisées hors ligne et indépendantes du processus de localisation.

Nous avons adopté la démarche suivante :

- Un premier sous-ensemble est utilisé comme base d'apprentissage : 10 positions sont utilisées pour estimer les valeurs des paramètres du modèle de propagation en fonction de l'emplacement de chaque landmark : il s'agit de la valeur du paramètre η de perte de trajet, des valeurs de l'atténuation de référence $PL(d_0)$ et de la variance du bruit σ^2. Nous pouvons aussi utiliser les positions des landmarks et les valeurs de RSSI mesurées de ces derniers pour estimer les valeurs des paramètres du modèle de propagation en fonction de l'emplacement de chaque landmark de l'architecture de localisation.

- Un deuxième sous-ensemble est considéré comme la base de test. Il est utilisé pour évaluer la précision des algorithmes. La source à localiser se localise elle-même sur ces positions.

La figure 4.20 présente une vue de dessus de la carte de nos testbeds lorsqu'on applique notre principe de base d'apprentissage et de base de test. Les cercles rouges illustrent les positions des landmarks. Les carrés jaunes à eux seuls constituent la petite base d'apprentissage nommée A : 10 positions sont utilisées pour apprendre les paramètres du modèle de propagation LNSM par rapport à l'emplacement de chaque landmark. Les carrés jaunes en plus des triangles violets représentent la grande base d'apprentissage nommée B : 25 positions sont utilisées pour apprendre les paramètres du modèle de propagation LNSM. Les astérisques bleus illustrent les bases de tests, les positions à localiser du nœud mobile aveugle.

Dans les sections qui suivent, nous faisons d'abord un rappel des algorithmes du maximum de vraisemblance et de l'algorithme de la trilateration. Nous présentons ensuite les algorithmes de localisation proposés. Nous comparons sur des plateformes expérimentales réelles nos algorithmes à ceux mentionnés de l'état de l'art.

Algorithmes de localisation à un saut basés sur la technique du RSSI

Nous présentons les heuristiques proposées pour corriger les incohérences observées. Nos algorithmes tiennent compte des effets de la propagation par trajets multiples et des valeurs aberrantes.

L'algorithme du maximum de vraisemblance

Le Maximum de Vraisemblance (i.e. MLE) est une approche statistique couramment utilisée pour localiser une position d'un nœud aveugle. Il se base sur la théorie de l'inférence statistique classique. Etant donné l'emplacement de chaque nœud de référence et la distance séparant chacun d'eux de la source à localiser, l'algorithme du maximum de vraisemblance maximise la probabilité de l'emplacement possible de la source à localiser en minimisant la variance de l'erreur estimée. Cependant, la performance de l'estimateur du maximum de vraisemblance est sensible au nombre de landmarks participants. Voici la fonction mathématique que nous avons utilisé pour estimer la position de la source à localiser :

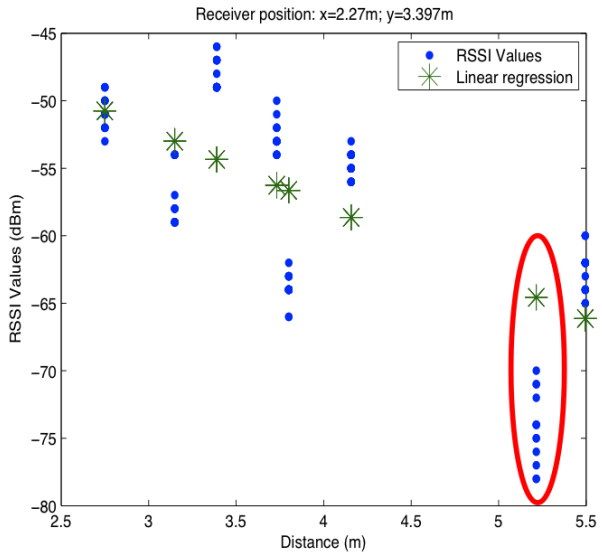
$$(\hat{x},\hat{y}) = \arg \min_{x\in(0,X),y\in(0,Y)} \sum_{\mathcal{L}=1}^{L} J_{\mathcal{L}} \log(\sigma_{\mathcal{L}}^2) + \sum_{\mathcal{L}=1}^{L} \sigma_{\mathcal{L}}^{-2} \sum_{j=1}^{J_{\mathcal{L}}} (A_{\mathcal{L},j} - \alpha_{\mathcal{L},0} - \alpha_{\mathcal{L},1} \log_{10} d_{\mathcal{L}}(x,y))^2 , \tag{6.2}$$

avec (x,y), le paramètre d'intérêt, autrement dit la position inconnue du nœud mobile aveugle en 2D ; $RSSI_{i,j}$ est l'atténuation j mesuré à la position inconnue du nœud mobile aveugle à partir du message envoyé par le landmark j. σ_i^2 est la variance du bruit associée au landmark i. J_i est le nombre total d'observations (de valeurs de RSSI mesurés) provenant du landmark i.

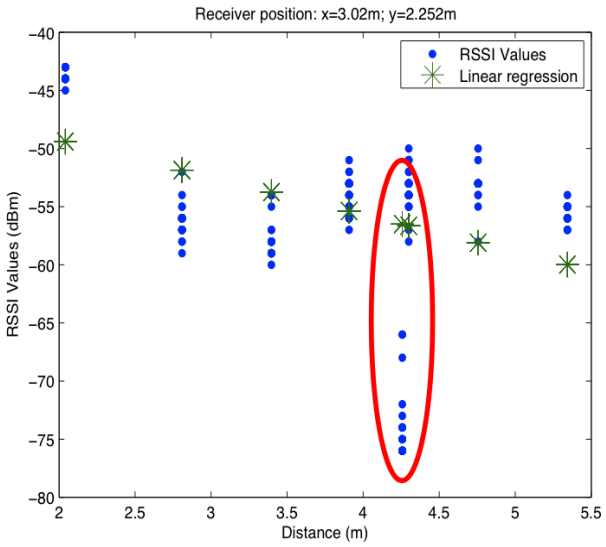
Grâce à cette équation présentée ci-dessus (cf. equation 6.2) et ne prenant en entrée que des valeurs de RSSI et les valeurs des paramètres du modèle de propagation, on détermine la position du nœud mobile aveugle avec l'algorithme du maximum de vraisemblance.

L'algorithme de la trilateration

Un autre algorithme fréquemment utilisé pour localiser la position d'un nœud mobile aveugle est la trilateration. C'est la technique de localisation utilisée par le GPS. La trilateration est une technique géométrique qui se base sur le principe d'intersection de cercles. Il utilise les propriétés géométriques de l'emplacement des points de référence pour estimer la position du nœud mobile aveugle. La distance séparant le nœud mobile aveugle de chaque point de référence est le rayon d'un cercle dont le centre est le point de référence qui a envoyé le message sur lequel la valeur de RSSI est mesurée. L'intersection de tous les points de référence donne la position du nœud mobile aveugle. Pour faire de la trilateration en environnement 2D, il faut deux points de référence et un nœud mobile aveugle. En environnement 3D, la trilateration nécessite 3 points de référence et un nœud mobile aveugle.



(a) 1ère position de test (#11) : x=2.27 m ; y=3.397 m



(b) 2eme position de test (#17) : x=3.02m ; y=2.252 m

FIGURE 6.15 – RSSI en fonction de la distance mesurés sur un nœud récepteur placé à deux positions différents (choisis de manière aléatoire) à partir des messages envoyés par 8 landmarks sur le testbed Rammus Rennes (#3)

L'algorithme de la localisation aléatoire

Nous avons aussi utilisé une technique de localisation aléatoire pour déterminer la contribution des mesures de RSSI. Pour cela, nous déterminons des coordonnées aléatoires suivant une distribution uniforme sur toute la taille du testbed en utilisant la méthode Monte Carlo. Cette dernière nous permet d' être sûr que la valeur aléatoire générée suit une loi recherché. Nous concernant, il s'agit de la loi uniforme.

Nous allons maintenant illustrer les algorithmes proposés pour prendre en compte les effets de propagation par trajets multiples et les valeurs aberrantes.

Localisation avec les algorithmes proposées

Nous avons représenté séparément les valeurs de RSSI mesurées sur deux positions différentes d'un nœud mobile aveugle à partir des messages reçus des 8 landmarks du testbed Rammus Rennes. Ces positions du nœud mobile aveugle sont choisies au hasard. Les résultats présentés à la figure 6.15 mettent en évidence que dans la plupart des cas, les valeurs de RSSI sont affectées de bruit et la variance du RSSI reste stable. Cependant, pour quelques liens de communication illustrés par des cercles rouges, la variance devient très grande et introduit un biais.

Ces résultats présentés ci-dessus et ceux des études expérimentales mentionnés plus haut et détaillés dans le chapitre 3, ont démontré que certaines valeurs de RSSI peuvent être affectées de grandes dispersions comparées aux autres. Autrement dit, un landmark parmi plusieurs peut être affecté d'un comportement anormal.

Nos premiers algorithmes de localisation proposés tiennent compte de la possibilité qu'un landmark en mauvaise condition de propagation par rapport à la position d'un récepteur peut être

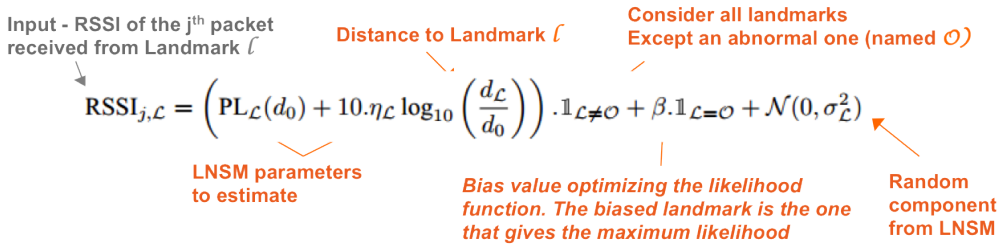


FIGURE 6.16 – Equation de Biased Log-Normal Shadowing Model (B-LNM)

affecté des effets de la propagation par trajet multiples induisant un comportement anormal. Les signaux reçus de ce landmark sont entachés de bruit. Les valeurs de RSSI mesurés des messages reçus de ce landmark anormal sont corrompues et de ce fait les effets de ces valeurs corrompues sont minimisés dans le processus de localisation. Nous proposons un modèle d'observation biaisée nommé : Biased-log-normal shadowing (B-LNM). Nous appliquons ensuite l'estimateur du maximum de vraisemblance pour permettre la localisation.

B-LNM ou le modèle de réduction des effets des valeurs de RSSI corrompues mesurées à partir des messages provenant d'un landmark anormal

B-LNM considère certaines valeurs de RSSI comme étant corrompues comparées aux autres. Ensuite ce modèle réduit les effets de ces valeurs corrompues provenant de landmarks anormaux suivant une re-paramétrisation du modèle log-normale shadowing qui conduit à ce nouveau modèle d'observation (cf. figure 6.16) :

$\mathrm{RSSI}_{j,\mathcal{L}}$ est la valeur du RSSI N$^{\underline{o}}$ j mesuré du message envoyé par le landmark $\mathcal{L}$ situé à une distance inconnue $d_{\mathcal{L}}$. Nous retrouvons les paramètres du modèle de propagation log-normal shadowing : $\mathrm{PL}_{\mathcal{L}}(d_0)$, $\eta_{\mathcal{L}}$ et $\mathcal{N}(0, \sigma^2_{\mathcal{L}})$.

L'équation de B-LNM prend en entrée toutes les valeurs de RSSI mesurées de tous les landmarks et calcule instantanément les paramètres du modèle de propagation dépendant de l'emplacement de chaque landmark. B-LNM définit ensuite le modèle statistique pour réduire les effets des valeurs anormales de RSSI mesurées des messages provenant d'un landmark anormal. Le paramètre β de l'équation de B-LNM (cf. figure 6.16) est un entier à optimiser, c'est l'indice du modèle sélectionné. Ensuite nous appliquons à B-LNM l'estimateur du maximum de vraisemblance pour localiser un nœud mobile aveugle en environnement interne.

Algorithme de localisation Biased-Maximum Likelihood (B-MLE)

En appliquant l'estimateur du maximum de vraisemblance à B-LNM, cela conduit aux maximisations suivantes :

La fonction de vraisemblance associée à tout landmark l et la fonction de vraisemblance associée au landmark $\mathcal{O}$ biaisé. Grâce à ces maximisations on détermine l'indice du modèle sélectionné, la position du nœud mobile aveugle, mais également la valeur du biais et le landmark biaisé s'il existe. En effet le landmark biaisé est le landmark qui étant exclu conduit à une valeur de

vraisemblance dont le maximum est le plus grand.

Grâce à la proposition d'un algorithme de localisation simple nommé B-MLE, nous localisons un nœud mobile aveugle tout en prenant en compte les effets de la propagation par trajets multiples. Cet algorithme se compose d'un modèle alternatif au modèle log-normal shadowing sur lequel est appliqué l'estimateur du maximum de vraisemblance. Cet algorithme de localisation n'est pas plus complexe que la plupart des algorithmes de l'état de l'art se basant sur la borne de *Cramer Rao* par exemple [28], [159], [31]. Outre les opérations de multiplications, divisions et d'additions, le log de la vraisemblance qui intervient dans l'équation de B-MLE peut être facilement supporté par les capteurs grâce au développement limité de la fonction logarithme.

Versions étendu de B-MLE : xB-MLE

Pour examiner les performances de l'algorithme de localisation proposé, nous avons étendu B-MLE pour prendre en compte le cas où plusieurs landmarks peuvent être affectés par un comportement anormal. Plus précisément, nous confrontons les cas où un (B-MLE), deux (2B-MLE) ou trois (3B-MLE) landmarks sont anormaux à l'estimateur classique du maximum appliqué au modèle log-normal shadowing (MLE non biaisé) sur les trois différents testbeds.

Présentée dans l'équation 6.3, la fonction de vraisemblance lorsque deux landmarks sont considérés comme biaisés se déduit aisément de la fonction de vraisemblance de B-MLE (i.e. un et un seul landmark peut être anormal). Le même principe s'applique lorsque trois landmarks sont considérés comme anormaux.

$$L_{\mathcal{O}}(x, y, \beta_1, \beta_2) = \sum_{\mathcal{L}=1; \mathcal{L}\neq\mathcal{O}_1; \mathcal{L}\neq\mathcal{O}_2}^{K} L_{\mathcal{L}}(x, y) + L_{\mathcal{O}_1}(\beta_1) + L_{\mathcal{O}_2}(\beta_2) \quad (6.3)$$

Nous avons également évalué une deuxième approche qui n'a pas besoin d'examiner tous les sous-ensembles possibles de landmarks anormaux. Au lieu de cela, cette approche sélectionne efficacement les landmarks normaux sur la base d'un seuil dépendant de la valeur du bais et défini dans l'équation 6.4.

$$\mathcal{B}_{\mathcal{L}} = \left| \frac{\beta_{\mathcal{L}} - \left(\mathrm{PL}_{\mathcal{L}}(d_0) + 10\eta_{\mathcal{L}} \log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right)\right)}{\mathrm{PL}_{\mathcal{L}}(d_0) + 10\eta_{\mathcal{L}} \log_{10}\left(\frac{d_{\mathcal{L}}}{d_0}\right)} \right| \quad (6.4)$$

Illustrée sur la figure 6.17, la valeur du seuil de cet algorithme de localisation nommé B-MLEC est choisie de façon empirique. Nous avons réalisé des expériences sur nos différents testbeds en confrontant les résultats obtenus lorsque différentes valeurs de seuils sont appliquées sur l'intervalle $]0; 0.7]$. Le seuil de valeur $\gamma = 0.05$ est la valeur optimale en terme de précision de localisation. En effet la valeur de seuil $\gamma = 0.05$ obtient la plus petite erreur de localisation pour tous les testbeds.

Nous allons maintenant présenter les résultats et comparer en environnements réels les précisions de localisation de B-MLE, MLE non biaisé et la trilateration sur les trois testbeds présentés à savoir Lincs, Barrault et Rammus.

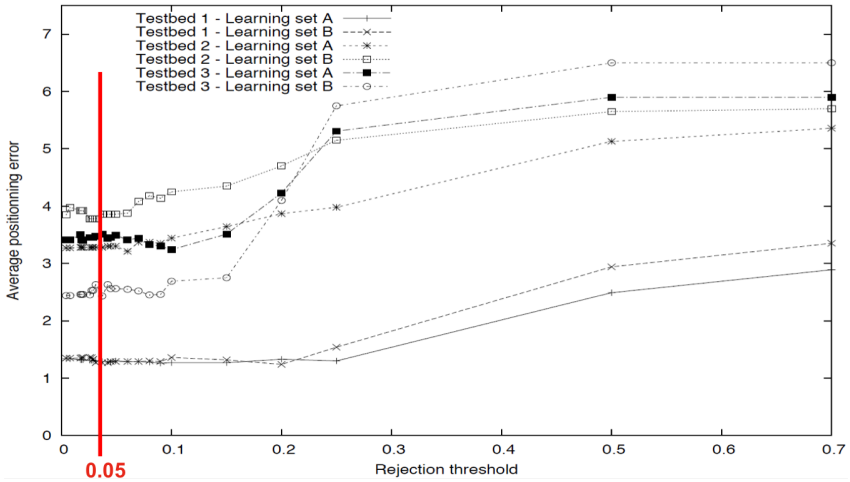


FIGURE 6.17 – Erreurs de localisation obtenues par B-MLEC dans les différents testbeds pour différentes valeurs du seuil $\mathcal{B}_{\mathcal{L}}$

Résultats expérimentaux

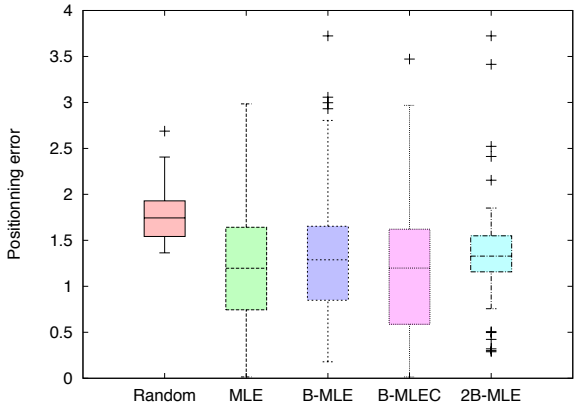
Nous avons comparé dans nos différents testbeds, la performance des algorithmes proposés (B-MLE, xB-MLE (2B-MLE, 3B-MLE) et B-MLEC) prenant en compte les effets de la propagation par trajets multiple à l'algorithme de localisation basé sur le maximum de vraisemblance (MLE) appliqué au modèle log-normale shadowing (MLE non biaisé). Nous avons également comparé nos algorithmes à la technique de la localisation aléatoire pour, rappelons le déterminer la contribution des mesures de RSSI.

Les résultats sont présentés sous forme de box-plots ou boites à moustaches qui sont des outils statistiques permettant de représenter la distribution des données de sorte que les valeurs aberrantes sont tracées à l'extérieur du graphique de la boite.

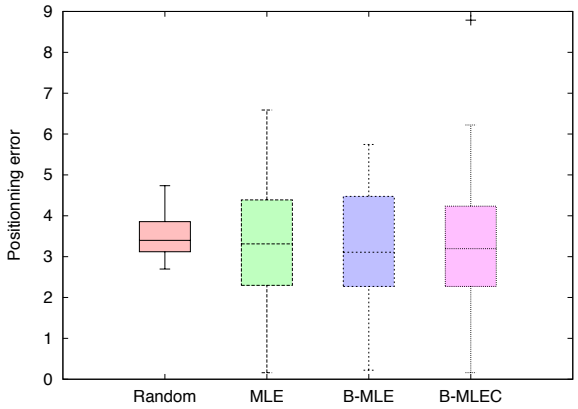
Si nous nous concentrons sur le testbed Lincs, le plus petit testbed, considérer plus d'un landmark comme étant anormal (2B-MLE) n'améliore pas la précision de localisation.

Pour le plus petit testbed, Lincs, il n'y a pas eu d'amélioration de la précision en considérant deux landmarks comme étant biaisés (2B-MLE) ; en effet, les résultats présentés sur la figure 6.18(a) montrent que la valeur médiane de 1.4 m obtenue par 2B-MLE n'est pas plus petite que celle obtenue par B-MLE qui est de 1.39 m. Par contre, lorsque la taille de la base d'apprentissage augmente, la précision de localisation n'est pas améliorée. Les résultats obtenus par B-MLEC, l'algorithme basé sur le seuil, sont meilleurs comparés à ceux obtenus par B-MLE et encore meilleurs que 2B-MLE. Toutefois nous remarquons que les valeurs aberrantes disparaissent avec l'augmentation de la taille de la base d'apprentissage (cf. figure 6.19(a)).

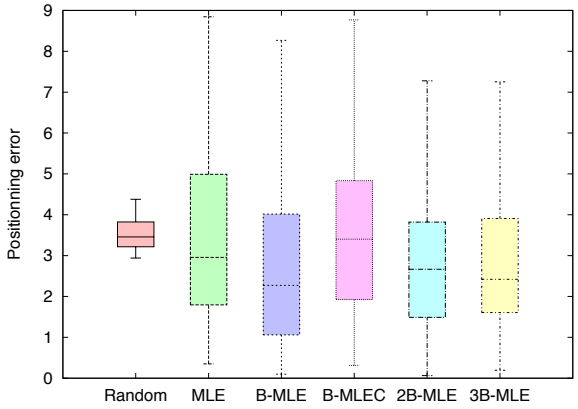
Les résultats obtenus pour le plus grand testbed, Rammus, ne démontrent pas une amélioration de la précision de localisation. Par exemple, pour la petite base d'apprentissage (i.e. A) (cf. figure 6.18(c)), la valeur médiane obtenue par 2B-MLE (2.67 m) n'est pas meilleure que celle obtenue par B-MLE (2.27 m). De plus, B-MLE obtient une meilleure performance de localisation comparée à 3B-MLE qui a obtenu une valeur médiane de 2.42 m. Lorsque nous comparons les valeurs moyennes, les résultats sont comparables pour les trois algorithmes. Par contre, à la différence du testbed Lincs, les résultats obtenus par B-MLEC pour le testbed Rammus ne sont pas meilleurs que ceux obtenus par B-MLE. La valeur médiane de B-MLEC 3.49 m ou encore la



(a) Testbed 1 (Lincs) - base d'apprentissage A (10 positions)

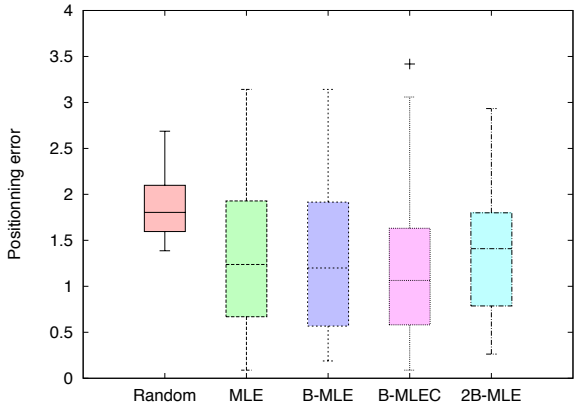


(b) Testbed 2 (barrault) - base d'apprentissage A (10 positions)

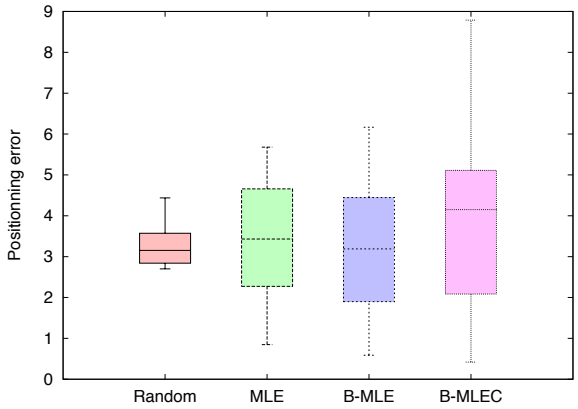


(c) Testbed 3 (Rammus) - base d'apprentissage A (10 positions)

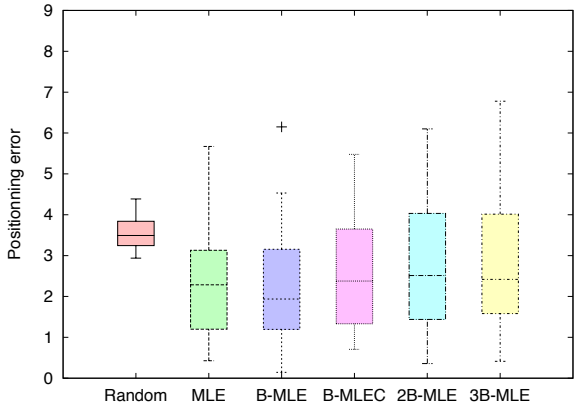
FIGURE 6.18 – Box-plots des erreurs de localisation obtenues par les différents algorithmes dans les différents testbeds pour la base d'apprentissage A (10 positions)



(a) Testbed 1 (Lincs) - base d'apprentissage B (25 positions)



(b) Testbed 2 (Barrault) - base d'apprentissage B (25 positions)



(c) Testbed 3 (Rammus) - base d'apprentissage B (25 positions)

FIGURE 6.19 – Box-plots des erreurs de localisation obtenues par les différents algorithmes dans les différents testbeds pour la base d'apprentissage B (25 positions)

valeur moyenne 3.40 m, sont respectivement supérieures à celles obtenues par B-MLE : 2.27 m comme valeur médiane et 2.73 m comme valeur moyenne. Il en résulte que diminuer le nombre de landmarks peut être rentable. L'estimation des paramètres du modèle de propagation s'effectuée plus rapidement. Ces résultats se confirment lorsque la taille de la base d'apprentissage augmente (cf. figure 6.19(c)). B-MLE obtient une meilleure performance de localisation comparée à B-MLEC. Une valeur médiane de 1.94 m et une valeur moyenne de 2.20 m obtenues par B-MLE sont meilleures en terme d'erreur de localisation comparés à 2.28 m de valeur médiane et 2.38 m de valeur moyenne obtenues par MLE non baisé. Ces performances de localisation sont encore meilleures que celles obtenues par B-MLEC : 2.38 m en valeur médiane et 2.56 m en valeur moyenne contre 2.42 m et 2.51 m en valeur médiane et 2.72 m et 2.73 m en valeur moyenne obtenues respectivement par 3B-MLE et 2B-MLE.

Outre le fait que B-MLE obtient une meilleure performance de localisation sur le testbed Rammus, ces résultats confirment aussi que le RSSI renseigne encore sur la distance, même si c'est peu car la localisation aléatoire obtient de mauvaises performances de localisation sur le testbed Rammus et Lincs comparées aux algorithmes que nous avons proposés quelle que soit la taille de la base d'apprentissage. De plus, l'ensemble des résultats montre une meilleure performance de localisation pour notre algorithme B-MLE comparés à MLE non biaisé quels que soient le testbed et la taille de la base d'apprentissage.

Les résultats globaux montrent que B-MLE améliore la précision de localisation comparée à MLE non biaisé. Nous avons aussi comparé les résultats individuels de localisation obtenus par B-MLE et MLE non biaisé sur le testbed Rammus. Chaque colonne du graphique de la figure 6.20 représente un point de mesure et plus la mesure est basse, plus précise est la méthode résultante. Nous avons remarqué :

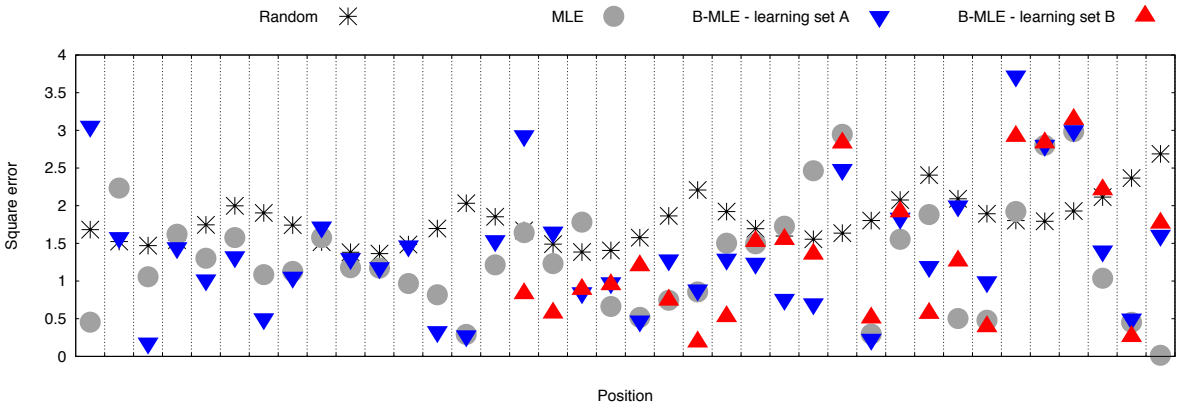
- Considéré un landmark comme étant anormal améliore significativement la précision de localisation sur le testbed Rammus, ceci n'est pas le cas sur le testbed Lincs. L'algorithme de sélection de landmarks anormaux basé sur le seuil (i.e. B-MLEC) obtient une meilleure performance de localisation.

- Les positions pour lesquelles B-MLE obtient une mauvaise performance sont souvent les mêmes quelle que soit la taille de la base d'apprentissage.

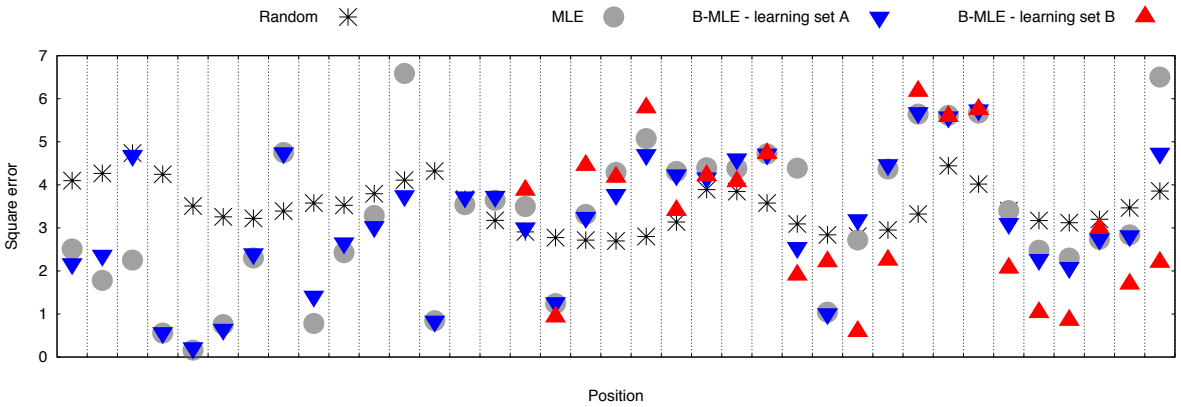
- La taille de la base d'apprentissage joue un rôle. Plus la base d'apprentissage est grande et plus les valeurs aberrantes disparaissent (petit testbed Lincs).

- B-MLE est capable de prendre en compte les effets de la propagation par trajets multiples. En effet, avec B-MLE la précision de localisation a été significativement améliorée sur la majeure partie des positions individuelles par rapport à MLE non biaisé (cf. figure 6.20). Les valeurs de RSSI mesurées sur ces positions étaient affectées de larges dispersions à cause des effets de la propagation par trajets multiples et B-MLE a été capable d'estimer avec une bonne précision ces positions (illustrées par les traits pleins en couleur verte).

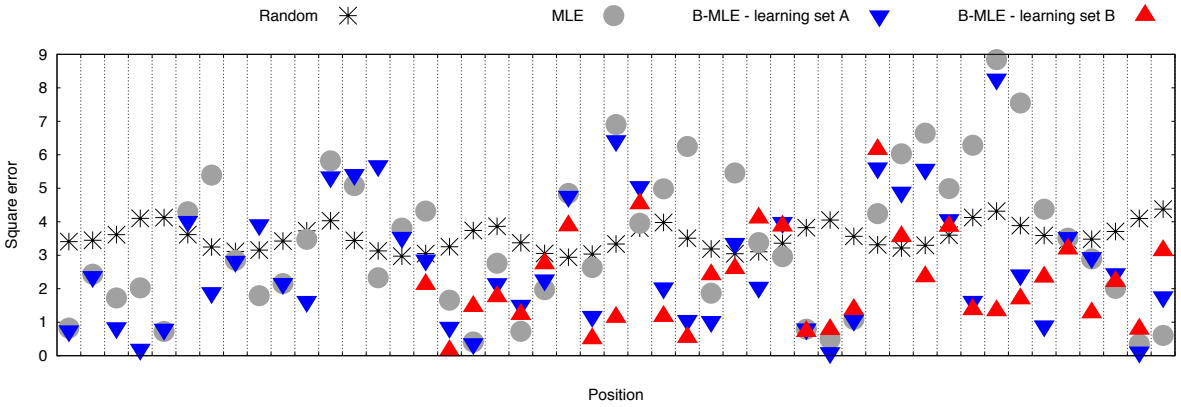
Nous avons prouvé qu'avec l'utilisation des seules valeurs de RSSI et y appliquer des traitements statistiques (B-MLE), il est possible d'améliorer la précision de localisation en environnement interne de 40% comparé au MLE non baisé (cf. tableau 6.1 6.3).



(a) Testbed #1



(b) Testbed #2



(c) Testbed #3

FIGURE 6.20 – Erreurs individuels de localisation pour la technique de localisation aléatoire, B-MLE et MLE non biaisé. Ces résultats de localisation relative sont obtenus pour les trois testbeds et en considérant deux bases d'apprentissage

TABLE 6.1 – Performance de localisation des différents algorithmes comparés sur le testbed #1

	Database	mean error	Median	STD
Random	A	1.79 m	1.74 m	0.31
MLE	A	1.29 m	1.20 m	1.29
B-MLE	A	1.39 m	1.29 m	0.85
2B-MLE	A	1.40 m	1.33 m	0.74
B-MLEC	A	1.30 m	1.20 m	0.85
Random	B	1.87 m	1.80 m	0.33
MLE	B	1.36 m	1.24 m	1.20
B-MLE	B	1.34 m	1.20 m	0.92
2B-MLE	B	1.39 m	1.41 m	0.75
B-MLEC	B	1.29 m	1.06 m	0.93

TABLE 6.2 – Performance de localisation des différents algorithmes comparés sur le testbed #2

	Database	mean error	Median	STD
Random	A	3.48 m	3.40 m	0.54
MLE	A	3.27 m	3.31 m	3.11
B-MLE	A	3.16 m	3.11 m	1.51
B-MLEC	A	3.30 m	3.20 m	1.76
Random	B	3.27 m	3.15 m	0.48
MLE	B	3.30 m	3.43 m	3.19
B-MLE	B	3.22 m	3.19 m	1.76
B-MLEC	B	3.86 m	4.01 m	2.38

TABLE 6.3 – Performance de localisation des différents algorithmes comparés sur le testbed #3

	Database	mean error	Median	STD
Random	A	3.53 m	3.46 m	0.39
MLE	A	3.41 m	2.96 m	2.27
B-MLE	A	2.73 m	2.27 m	1.93
2B-MLE	A	2.72 m	2.67 m	1.59
3B-MLE	A	2.72 m	2.42 m	1.68
B-MLEC	A	3.49 m	3.40 m	2.05
Random	B	3.55 m	3.49 m	0.40
MLE	B	2.38 m	2.28 m	1.94
B-MLE	B	2.20 m	1.94 m	1.40
2B-MLE	B	2.73 m	2.51 m	1.49
3B-MLE	B	2.72 m	2.42 m	1.43
B-MLEC	B	2.56 m	2.38 m	1.40

Algorithmes distribués proposés

Dans les algorithmes précédents, un nœud ne peut qu'estimer sa position. Pour se projeter dans un scénario plus proche des applications des réseaux de capteurs et de plusieurs autres applications, nous avons proposé un algorithme distribué dans lequel chaque capteur est capable d'estimer et d'affiner sa position mais également celle de ses voisins pour des besoins de coopération.

Algorithme d'Approximation Stochatisque Distribué (en anglais Distributed Stochastique Approximation Localization Algorithm : DSAL)

L'algorithme de localisation distribué proposé est le DSAL pour algorithme d'Approximation Stochatisque Distribué. Dans la mesure où les valeurs de RSSI mesurées sont bruitées en environnement interne, l'algorithme distribué DSAL affine les positions initialement estimées par les nœuds d'un réseau de capteur sans fil en deux étapes en ligne : la première est une étape de calcul de descente du gradient distribué sur chaque capteur. Il s'agit de la minimisation de l'erreur quadratique entre la distance obtenue sur les valeurs de RSSI mesurées, bruitées et les valeurs de RSSI qu'on obtiendrait avec les vraies positions recherchées. Autrement dit chaque nœud capteur calcule une estimation provisoire de sa position et de celles de ses voisins, suivant cette formule stochastique :

$$(\boldsymbol{x}_{t+1}^{(i)}, \boldsymbol{y}_{t+1}^{(i)}) = (\boldsymbol{x}_{t}^{(i)}, \boldsymbol{y}_{t}^{(i)}) - \gamma_t \nabla f_i(\boldsymbol{x}_{t}^{(i)}, \boldsymbol{y}_{t}^{(i)})$$

avec γ_t est un pas décroissant : $\gamma_t = 1/\sqrt{t}$ et qui intervient à chaque itération de l'algorithme. Cette étape de calcul de la descente du gradient distribué se fait en local sur chaque nœud capteur.

La seconde est une étape de communication entre deux capteurs et utilise la version bidirectionnelle du protocole de communication Gossip [33] : chaque itération de l'algorithme représente une étape de communication aléatoire entre deux capteurs. Deux nœuds capteurs i et j choisis au hasard communiquent pour s'échanger leurs positions estimées afin de converger vers un consensus sur l'ensemble des positions des nœuds capteurs qui sont voisins communs aux capteurs i et j. Si t représente l'instant de temps discret, chaque itération t de l'algorithme comporte l'ensemble des positions estimées par un nœud capteur. Il est à noter que seul les positions des nœuds qui sont voisins communs des nœuds i et j seront mis à jour suivant cette équation ci-dessous, exécutée jusqu'à ce que le protocole converge :

$$\forall \ell \in \mathcal{N}_i \cap \mathcal{N}_j, \tag{6.5}$$

$$\begin{aligned}(\boldsymbol{x}_{t+1}^{(i)}(\ell), \boldsymbol{y}_{t+1}^{(i)}(\ell)) &= \frac{(\boldsymbol{x}_{t+\frac{1}{2}}^{(i)}(\ell), \boldsymbol{y}_{t+\frac{1}{2}}^{(i)}(\ell)) + (\boldsymbol{x}_{t+\frac{1}{2}}^{(i)}(\ell), \boldsymbol{y}_{t+\frac{1}{2}}^{(i)}(\ell))}{2} \\ (\boldsymbol{x}_{t+1}^{(j)}(\ell), \boldsymbol{y}_{t+1}^{(j)}(\ell)) &= (\boldsymbol{x}_{t+1}^{(i)}(\ell), \boldsymbol{y}_{t+1}^{(i)}(\ell))\end{aligned} \tag{6.6}$$

Sinon, les positions des nœuds non voisins communs de i et j sont pris en compte (maintenus) suivant l'équation suivante :

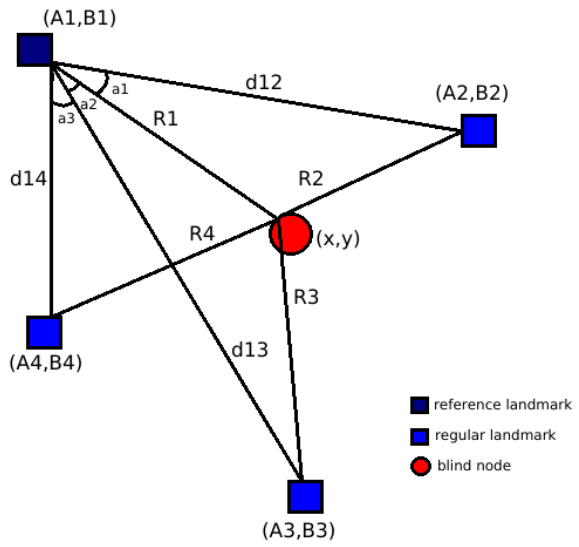


FIGURE 6.21 – Overview of the multilateration approach

$(\boldsymbol{x}_{t+1}^{(m)}(\ell), \boldsymbol{y}_{t+1}^{(m)}(\ell)) = (\boldsymbol{x}_{t+1}^{(m)}, \boldsymbol{y}_{t+1}^{(m)})$.

La précision de localisation de notre algorithme distribué est comparée à d'autres algorithmes de l'état de l'art : la multilatération, le multi dimensional scaling with modern majorization problem et le classical multi dimensional scaling.

Algorithme de la multilateration : état de l'art

La multilateration est une technologie basée sur l'analyse géométrique. C'est une extension de la trilateration utilisant le théorème du cosinus. Elle est utilisée pour localiser un nœud mobile aveugle entouré d'au moins 4 landmarks avec des positions connues $(x_i; y_i), i = 1, \ldots, 4$. A la différence de la trilateration où on mesure des distances, la multilateration (cf. figure 6.21) implique de mesurer la différence entre le carré des distances séparant le nœud mobile aveugle des landmarks.

Avec M landmarks, le théorème du cosinus est appliqué à n'importe lequel des $M-1$ triangles formés de trois sommets de capteurs composés entre les landmarks et le nœud mobile à localiser.

Le principe de la méthode de multilateration est de construire un système linéaire de $M-1$ équations à deux inconnues (coordonnées du nœud mobile à localiser $(x_{bn}; y_{bn})$) entre le nœud mobile à localiser, un landmark choisi comme référence et le reste des landmarks comme suit :

$$
\begin{aligned}
&d_1^2 + d_{12}^2 - 2\left((x_{BN} - x_1)(x_2 - x_1) + (y_{BN} - y_1)(y_2 - y_1)\right) = d_2^2 \\
&d_1^2 + d_{13}^2 - 2\left((x_{BN} - x_1)(x_3 - x_1) + (y_{BN} - y_1)(y_3 - y_1)\right) = d_3^2 \\
&d_1^2 + d_{14}^2 - 2\left((x_{BN} - x_1)(x_4 - x_1) + (y_{BN} - y_1)(y_4 - y_1)\right) = d_4^2 \\
&\vdots \\
&d_1^2 + d_{1M}^2 - 2\left((x_{BN} - x_1)(x_M - x_1) + (y_{BN} - y_1)(y_M - y_1)\right) = d_M^2
\end{aligned}
$$

Nous réécrivons l'expression ci-dessus sous forme d'une matrice afin d'avoir un système linéaire $\{d_i\}_{i=1,\ldots,M}$ formé des distances entre le nœud mobile aveugle et les M landmarks :

$$\boldsymbol{b} = A\boldsymbol{\theta} \quad \text{avec :} \quad \boldsymbol{b} = \boldsymbol{d} - \boldsymbol{\delta} = \frac{1}{2}\begin{bmatrix} d_{12}^2 - (d_2^2 - d_1^2) \\ \vdots \\ d_{1M}^2 - (d_M^2 - d_1^2) \end{bmatrix}$$

$$A = \begin{bmatrix} (x_2 - x_1) & (y_2 - y_1) \\ \vdots & \vdots \\ (x_M - x_1) & (y_M - y_1) \end{bmatrix} \quad \text{et} \quad \boldsymbol{\theta} = (x_{BN}, -x_1\ ,\ y_{BN}, -y_1)^t = \boldsymbol{z} - (x_1, y_1)^t \tag{6.7}$$

La solution du système d'équations, d'où les coordonnées inconnues du nœud mobile à localiser $(x_bn; y_bn)$ est obtenue en résolvant une optimisation linéaire. Il s'agit de la minimisation de l'erreur quadratique moyenne associée aux mesures de distance entre le nœud mobile à localiser et les landmarks du réseau, y compris le landmark de référence. Avec la multilateration, plusieurs nœuds aveugles peuvent se localiser en même temps.

Les algorithmes de localisation présentés ci-dessus sont également comparés aux approches de localisation basées sur le Multidimensional Scaling (MDS).

Algorithme de localisation basé sur le Multidimensional Scaling (MDS) : état de l'art

Le Multidimensional Scaling (MDS) est une technique d'analyse de données par le biais de laquelle on peut déduire les principales dimensions liées à la configuration spatiale des objets lorsque des données disponibles telles que les mesures de RSSI entre les nœuds permettent de renseigner sur la similitude générale ou la différence entre ces objets mesurés. La matrice proximité nécessaire avec MDS contient les carrés des distances entre tous les capteurs

Les données recueillies et nécessaires pour l'analyse MDS et permettant de renseigner la similarité ou la différence entre les objets sont représentées dans une matrice appelé matrice de proximité. Grâce à la matrice de proximité, on peut déduire les coordonnées de l'ensemble des nœuds du réseau soit en déterminant le sous espace principal lié à la matrice, soit en résolvant un problème de minimisation.

Dans la mesure où les distances entre les nœuds ne sont pas connues, nous supposons le modèle suivant : un ensemble N de nœuds aveugles formant un graphe connexe dont les distances sont estimées à partir des valeurs de RSSI mesurées sur chaque nœud aveugle et en utilisant l'estimateur sans biais défini dans l'équation $\hat{d}_{square}$:

$$\hat{d}_{square} = 10^{\frac{-\overline{\boldsymbol{R}} - \mathrm{PL}(d0)}{5\eta}} K^{-2} = d^2\, 10^{\frac{2\xi}{10\eta}}\, 10^{-\frac{\sigma^2 \ln 10}{T(10\eta)^2}} \quad \Longrightarrow \quad E[\hat{d}_{square}] = d^2 \tag{6.8}$$

T étant le nombre de valeurs de RSSI mesurées d'un nœud émetteur voisin et représentées dans un vecteur $\boldsymbol{R} = (R(1), \ldots, R(T))$. $\{\mathrm{PL}(d_0), \eta, \sigma^2\}$ sont les paramètres du modèle de propagation log-normal shadowing [22] définit dans l'équation (6.1). ξ est la variable aléatoire gaussienne de moyenne nulle et de variance σ^2.

Avec MDS, le nœud central calcule toutes les $N(N-1)/2$ distances entre chaque pair de nœud i et j, crée la matrice de distance résultante, fait le « double centering »et ensuite détermine les

positions relatives de l'ensemble des nœuds du réseau suivant deux méthodes différentes et en se basant sur la positions relatives de ses voisins. De ce fait les coordonnées de référence sont le barycentre du réseau formé par l'ensemble des nœuds du réseau. Le « double centering », est le mécanisme permettant de considérer le barycentre du réseau comme les coordonnées de référence. On part d'une matrice en fonction des positions absolues pour estimer la matrice de carré des distances suivant un référentiel relatif (barycentre du réseau).

Borg *et al.* ont proposé deux manières différentes de résoudre le problème de la localisation d'un ensemble N de nœuds capteurs aveugles avec des positions inconnues lorsque des données sont disponibles sous la forme d'une matrice de distance carrée et symétrique D_{square}. Les éléments $D_{square(i,j)_{\forall i,j}}$ de la matrice D_{square} correspondent à l'ensemble $N(N-1)/2$ des pairs de distances carrées entre les N nœuds du réseau. On note $\boldsymbol{Z}$, la matrice N contenant les positions de chaque nœud capteur du réseau et dont les éléments sont de dimension $N \times 2$. Les positions relatives $\boldsymbol{Z}$ sont obtenues suivant deux méthodes. Ces méthodes présentent les différentes manières de résoudre la matrice D_{square} en fonction de la matrice $\boldsymbol{Z}$:

Analyse de la composante principale (Principal Component Analysis (PCA))

Avec la méthode MDS d'analyse en composante principale, le nœud central détermine le sous espace principal lié à la matrice de distance. Autrement dit, il réalise une décomposition en vecteurs propres et en valeurs propres de la matrice de distance. Ces vecteurs propres et valeurs propres conduisent aux positions relatives recherchées des nœuds du réseau suivant ces équations :

$$\frac{-1}{2} J_{\perp}\, D_{square}\, J_{\perp} = \frac{-1}{2} J_{\perp}(c\mathbf{1}^t + \mathbf{1}c^t - 2\boldsymbol{Z}\boldsymbol{Z}^t)J_{\perp} = J_{\perp}\boldsymbol{Z}\boldsymbol{Z}^t J_{\perp} = \boldsymbol{X}\boldsymbol{X}^t \tag{6.9}$$

et

$$\boldsymbol{X}\boldsymbol{X}^t = V\Gamma V^t = V\sqrt{\Lambda}\sqrt{\Lambda}V^t \tag{6.10}$$

$$\text{alors} \quad : \quad \boldsymbol{X} = V\sqrt{\Lambda} = [\sqrt{\lambda_1}\boldsymbol{V}_1, \sqrt{\lambda_2}\boldsymbol{V}_2] \tag{6.11}$$

MDS moderne (SMACOF)

Au lieu d'utiliser la décomposition en vecteurs et valeurs propres, MDS moderne résout un problème de minimisation d'une fonction nommée « Raw Stress »exprimant l'erreur quadratique moyenne entre les distances réelles entre chaque pair de nœud i et j. $d_{i,j} = \sqrt{(z_1(i) - z_1(j))^2 + (z_2(i) - z_2(j))^2}$ et les distances estimées $\{\hat{d}_{i,j}\}_{\forall i,j}$ par l'intermédiaire des valeurs RSSI comme suit :

$$\hat{\boldsymbol{X}} = \min_{\boldsymbol{X}} \sigma_r(\boldsymbol{X}) = \min_{\boldsymbol{X}} \sum_{j>i} w_{ij}(\hat{d}_{i,j} - d_{i,j}(\boldsymbol{X}))^2 \tag{6.12}$$

$$\text{avec} : d_{i,j}(\boldsymbol{X}) = \sqrt{(X_1(i) - X_1(j))^2 + (X_2(i) - X_2(j))^2}$$

Utilisant PCA ou MDS moderne, le nœud central affiche en sortie une carte contenant les positions relatives de tous les nœuds du réseau. Cette carte peut être convertie suivant un référentiel absolu avec l'utilisation de landmarks.

Nous allons maintenant comparer ces algorithmes de l'état de l'art, à savoir les approches basées sur MDS, la multilateration avec notre proposition (i.e. DSAL).

Formule d'erreur de distance utilisée

Nous rappelons l'expression de l'erreur d'estimation moyenne de localisation (E_m) utilisée pour évaluer la performance des mécanismes de localisation en simulation et avec des données réelles des capteurs.

Si nous considérons un ensemble de N positions inconnues à localiser, $\{\boldsymbol{X}(i)\}_{i=1,\ldots,N}$ où chaque $\boldsymbol{X}(i)$ correspond à la position estimée en 2D du nœud de capteur i, l'erreur moyenne de localisation de l'ensemble des positions estimées $\{\hat{\boldsymbol{X}}(i)\}_{i=1,\ldots,N}$ est définie par E_m :

$$E_m = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\boldsymbol{X}}(i) - \boldsymbol{X}(i)\| \tag{6.13}$$

Dans l'équation 6.13 nous avons des unités de distance en mètre m pour nos expériences. Pour obtenir une valeur qualitative et absolue en pourcentage de l'erreur de localisation moyenne, l'expression (6.13) peut être normalisée par un facteur qui prend en compte les dimensions du testbed dans lequel évolue le réseau de capteurs sans fil. Par exemple, pour les expériences avec un réseau de capteurs sur un rectangle de dimensions $Pm \times Qm$, la valeur de la distance d est normalisée par la diagonale du rectangle, pour obtenir une estimation de l'erreur moyenne normalisée (NME) comme suit :

$$\text{NME} = \frac{1}{\sqrt{P^2+Q^2}}\frac{1}{N}\sum_{i=1}^{N} \|\hat{\boldsymbol{X}}(i) - \boldsymbol{X}(i)\| \tag{6.14}$$

Si l'erreur moyenne normalisée (NME) est calculée pour chaque position i, elle peut être définie par $\frac{1}{\sqrt{P^2+Q^2}}\|\hat{\boldsymbol{X}}(i) - \boldsymbol{X}(i)\|$.

Comparaisons des algorithmes de localisation distribués en environnements réels

Les algorithmes ont d'abord été testés en simulation en générant 1000 réalisations de Monte Carlo sous Matlab avec une distribution log-normale. 12 nœuds capteurs aveugles sont déployés sur une surface de 10 m par 7 m.

Nous avons également comparé ces algorithmes sur la plateforme réelle de capteurs, SensLAB, Rennes avec 12 nœuds capteurs aveugles déployés sur une surface de 8 m par 5 m, 4 landmarks et 10 expériences indépendantes de l'algorithme DSAL.

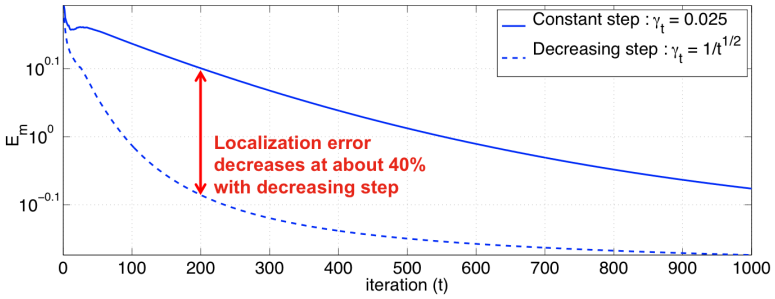


FIGURE 6.22 – Convergence de l'algorithme DSAL en simulation

Convergence de l'algorithme DSAL en simulation

Les résultats de simulation présentés sur la figure 6.22 montrent qu'une meilleure précision de l'algorithme DSAL est obtenue lorsqu'on considère un pas décroissant. En effet, en utilisant un pas décroissant au lieu d'un pas croissant, on passe d'une erreur de localisation de 1.25 m à en dessous de 0.79 m en 200 itérations. L'erreur de localisation est réduite de 40

Précision d'erreur de localisation de l'algorithme DSAL sous la présence de bruit

Nous avons étudié en simulation l'erreur de localisation obtenue par DSAL par rapport à la présence de bruit. Nous ne remarquons pas une augmentation significative de l'erreur de localisation lorsque la variance du bruit augmente. Par contre les résultats de localisation obtenus par la multilateration ou par les méthodes basées sur MDS démontrent une erreur de localisation qui augmente avec la variance du bruit.

Les résultats de localisation sont présentés sous forme d'erreur moyenne normalisée de localisation suivant les dimensions P et Q du testbed concerné comme dÃ©fini dans l'équation 6.14 :

En effet, avec DSAL, nous obtenons une erreur moyenne de localisation autour de 0.05 m lorsque la variance du bruit passe de $\sigma^2 = 2$ à $\sigma^2 = 10$. Par contre la multilateration obtient une erreur moyenne normalisée de localisation supérieure à 0.09 et 0.2 pour les mêmes variances du bruit. Les méthodes basées sur MDS quant à elles obtiennent une erreur moyenne normalisée de localisation comprise entre $0.05 \leq NME \leq 0.1$ lorsque $\sigma^2 = 2$. Si la variance du bruit augmente de $\sigma^2 = 10$, l'erreur moyenne de localisation est très supérieure à 0.1 ($NME \geq 0.1$). Ces résultats de simulation présentés dans le tableau 6.23 confirment que l'algorithme DSAL obtient une meilleure précision de localisation sous la présence de bruit. Cette performance est meilleure que celles obtenues par les algorithmes de la multilateration et ceux basés sur MDS.

La figure 6.24 confirme les résultats présentés ci-dessus tout en montrant les courbes de comparaisons des erreurs moyennes normalisées de localisation obtenues par les quatre algorithmes comparés.

Comparaison des algorithmes sur une plateforme réelle de capteurs

Les résultats de localisation obtenus durant les expériences en environnement réel sont présentés dans le tableau 6.4. Ils ne confirment pas ceux obtenus en simulation. Ce qui est normal car

	DSAL (DG)	Multilateration	MDS-based methods
$\sigma^2 = 2$	$NME < 0.05$	$NME \geq 0.09$	$0.05 \leq NME \leq 0.1$
$\sigma^2 = 10$	$NME = 0.05$	$NME \geq 0.2$	$NME \geq 0.1$

FIGURE 6.23 – Tableau de comparaison des résultats de localisation obtenus en simulation avec l'algorithme DSAL, de la multilateration (MC), et des methodes basées sur Multi Dimensional Scaling (Modern et MDS classique)

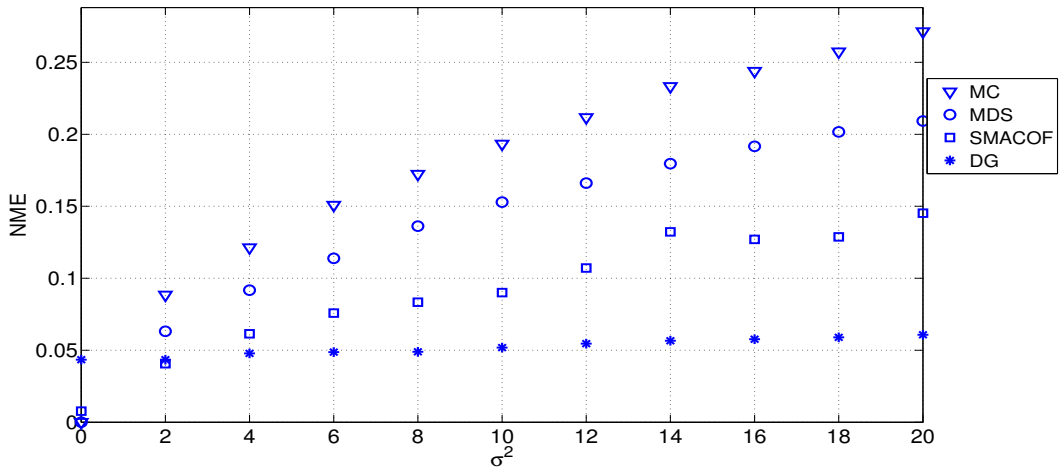


FIGURE 6.24 – Erreur moyenne normalisée (NME) estimée en simulation avec DSAL avec 1000 réalisations de Monte Carlo pour $N = 12$ positions inconnues de nœuds aveugles pour différentes valeurs de la variance σ^2 du bruit. Un pas constant $(\gamma_t = 0.025)_{t\geq 1}$ et une initialisation aléatoire sont considérés pour l'algorithme DSAL.

la simulation ne prend pas en compte tous les facteurs perturbateurs de la propagation en environnement réel.

TABLE 6.4 – Erreurs moyennes normalisée (NME) et absolues (E_m)) de localisation obtenues par 12 nœuds capteurs aveugles en éxecutant les différents algorithmes de localisation comparés sur la plateforme SensLAB Rennes.

Methodes	**EMN (%)**	$\boldsymbol{E_m}$ **(m)**
Multilateration (MC)	0.23	2.2
MDS classique	0.3	2.8
MDS moderne (SMACOF)	0.25	2.4
MLE	0.28	2.63
DG (pas décroissant et initialisation aléatoire)	0.14	1.3
DG (pas croissant et initialisation aléatoire)	0.08	0.77
DG (pas décroissant et initialisation avec MLE)	0.1	0.98
DG (pas croissant et initialisation avec MLE)	0.076	0.72

En effet une meilleure précision de localisation est obtenue avec l'algorithme DSAL lorsqu'un pas constant est considéré au lieu d'un pas décroissant (cf. figures 6.26). Et cette précision de localisation est encore meilleure lorsqu'on considère une initialisation avec le maximum de vraisemblance (MLE) (i.e. pas constant + initialisation avec MLE pour DSAL). Une erreur

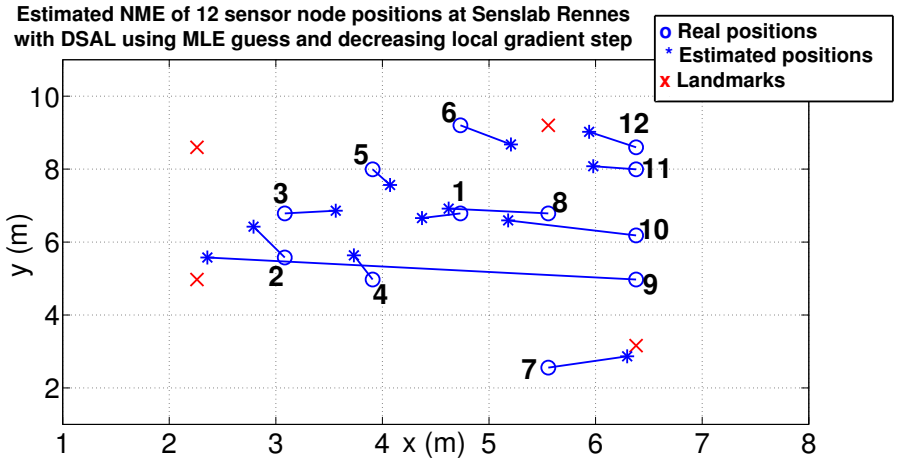
moyenne normalisée de 0.72 m est obtenue avec DSAL en considérant un pas constant et une initialisation avec MLE. Cette erreur moyenne de localisation est meilleure (i.e. plus petite) que 0.77 m obtenue lorsque l'algorithme DSAL est considèré avec un pas constant et une initialisation aléatoire. Ces résultats sont meilleurs que 0.98 m et 1.3 m obtenus lorsqu'on considère DSAL avec un pas décroissant et une initialisation avec MLE puis une initialisation aléatoire. Toutefois ces résultats d'erreur de localisation obtenus par DSAL sont meilleurs que ceux obtenus par le MDS classique : 2.8 m, encore meilleurs que ceux obtenus par MLE= 2.63 m, et MDS moderne : 2.4 m et la multilateration : 2.2 m. Les figures 6.25 et 6.26 confirment ces résultats présentés ci-dessus et montrent les performances de localisation de l'algorithmes DSAL sur la plateforme réelle de capteurs de SensLAB Rennes.

Les erreurs de localisations individuelles obtenues par les algorithmes distribués comparés

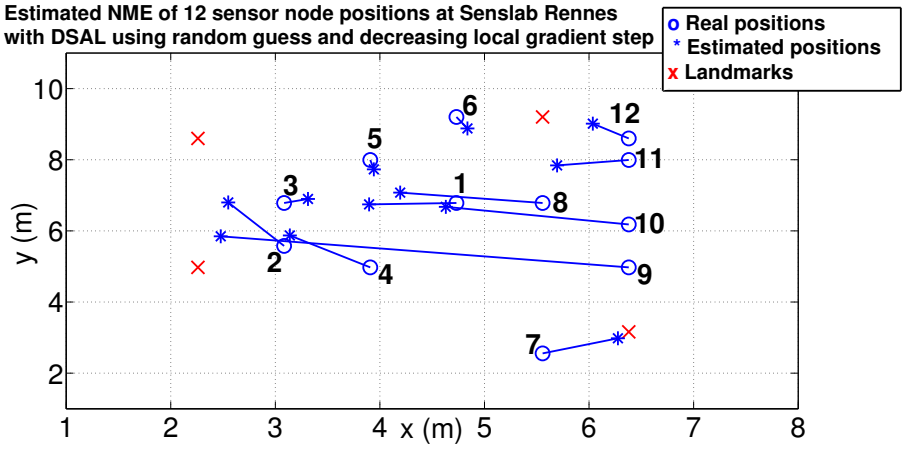
Nous avons aussi comparé les erreurs de localisations individuelles obtenues par les algorithmes DSAL, multilateration et ceux basés sur MDS (cf. tableau 6.5 et figure 6.27).

TABLE 6.5 – Valeurs de NME pour chacun des 12 nœuds aveugles à localiser pour différentes différents algorithmes de localisation et en considérant 10 expériences indépendentes sur la plateforme SensLAB Rennes.

	Simulated Matlab data				**Real SensLAB data**			
ID du nœud	**MC**	**MDS**	**SMACOF**	**DG**	**MC**	**MDS**	**SMACOF**	**DG**
214 / 1	0.15	0.13	0.093	0.051	0.13	0.23	0.19	0.02
176 / 2	0.21	0.17	0.13	0.056	0.2	0.29	0.17	0.064
178 / 3	0.20	0.16	0.13	0.064	0.05	0.05	0.27	0.067
193 / 4	0.19	0.15	0.12	0.056	0.21	0.35	0.23	0.048
198 / 5	0.23	0.16	0.11	0.053	0.2	0.02	0.15	0.066
218 / 6	0.34	0.22	0.11	0.054	0.45	0.14	0.28	0.03
225 / 7	0.41	0.27	0.15	0.071	0.22	0.59	0.51	0.06
232 / 8	0.17	0.16	0.1	0.051	0.21	0.12	0.24	0.088
247 / 9	0.23	0.21	0.13	0.06	0.2	0.47	0.063	0.35
249 / 10	0.21	0.20	0.12	0.054	0.25	0.29	0.41	0.07
252 / 11	0.28	0.23	0.15	0.052	0.25	0.32	0.35	0.032
253 / 12	0.33	0.25	0.15	0.055	0.35	0.12	0.18	0.012

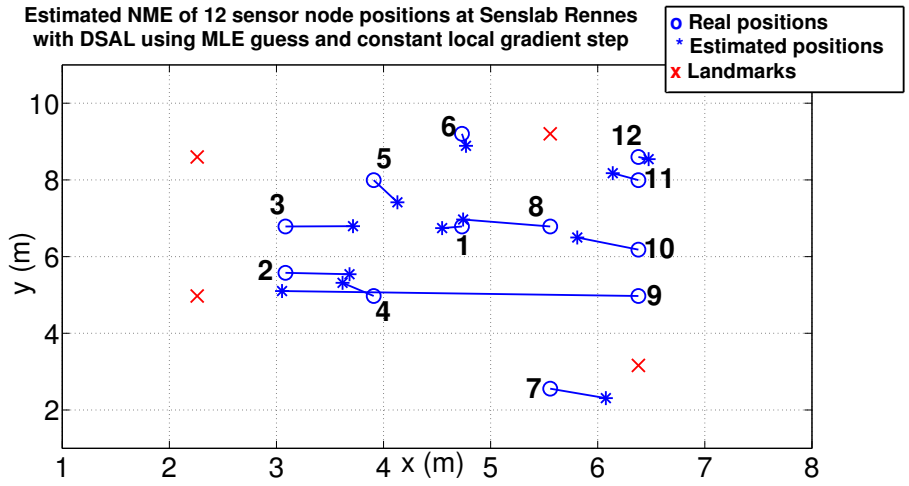


(a) Performance de localisation de l'algorithme DSAL en environnement réel : itération de DSAL avec un pas décroissant et une initialisation avec MLE$(\gamma_t = 1/\sqrt{t})_{t\geq 1}$.

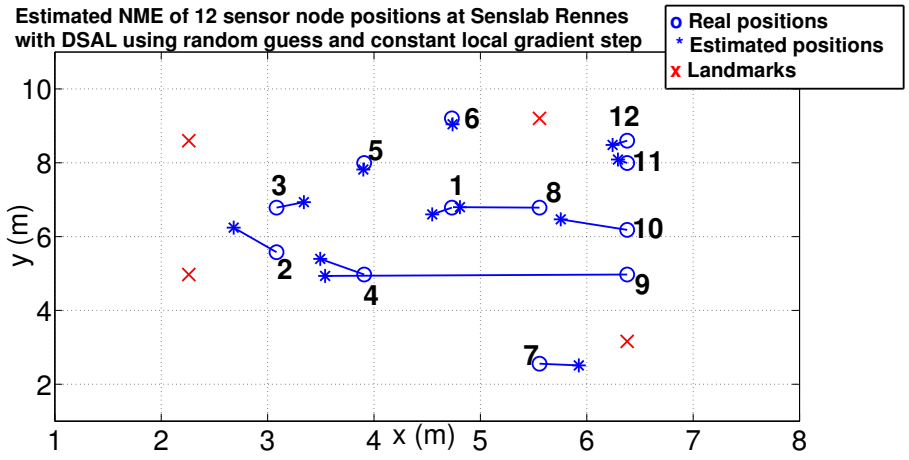


(b) Performance de localisation de l'algorithme DSAL en environnement réel : itération de DSAL avec un pas décroissant et une initialisation aléatoire$(\gamma_t = 1/\sqrt{t})_{t\geq 1}$.

FIGURE 6.25 – Résultats de localisation de $N = 12$ nœuds aveugles (présentés sous forme de carte) sur la plateforme SensLAB Rennes avec l'algorithme DSAL lorsqu'on considère un pas décroissant, une initialisation aléatoire et une initialisation avec MLE



(a) Performance de localisation de l'algorithme DSAL en environnement réel : itération de DSAL avec un pas constant et une initialisation avec MLE $(\gamma_t = 0.025)_{t\geq 1}$.



(b) Performance de localisation de l'algorithme DSAL en environnement réel : itération de DSAL avec un pas constant et une initialisation aléatoire $(\gamma_t = 1/\sqrt{t})_{t\geq 1}$.

FIGURE 6.26 – Résultats de localisation de $N = 12$ nœuds aveugles (présentés sous forme de carte) sur la plateforme SensLAB Rennes avec l'algorithme DSAL lorsqu'on considère un pas constant ; une initialisation aléatoire et une initialisation avec MLE

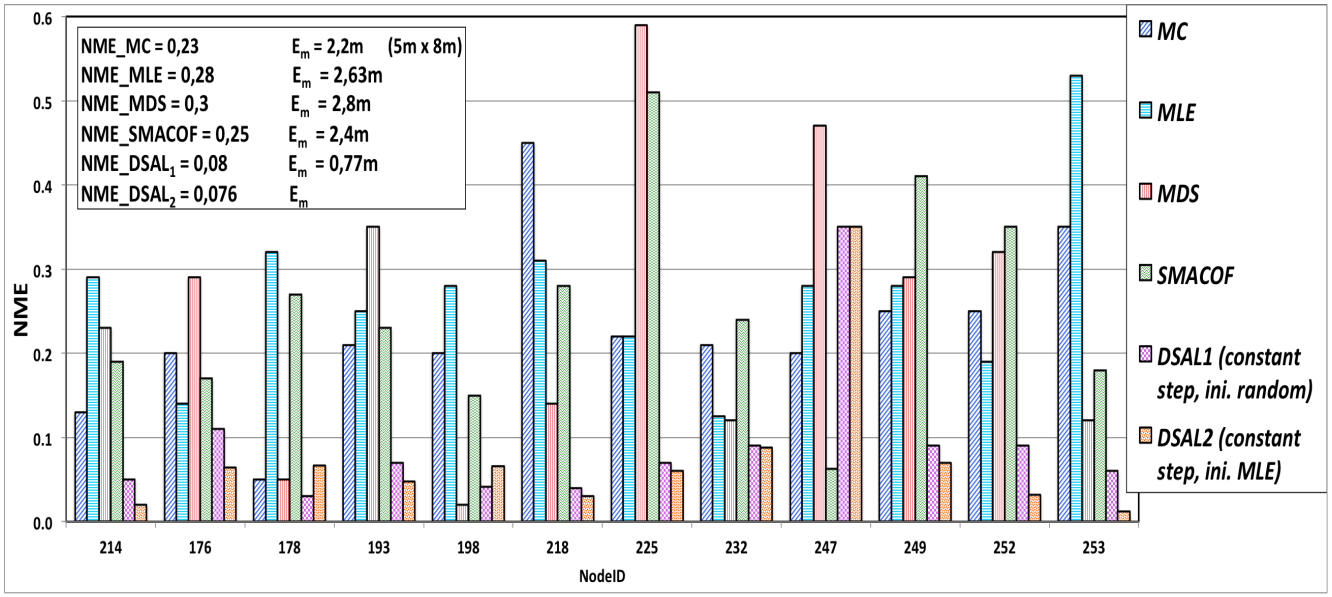


FIGURE 6.27 – Erreur moyenne normalisée (NME) de localisation obtenue par chaque nœud aveugle individuel pour les différents algorithmes sur la plateforme SensLAB Rennes. 10 expériences indépendantes sont réalisées. Le réseau est composé de 12 nœuds aveugles et 4 landmarks.

L'algorithme de la multilateration obtient une mauvaise performance de localisation pour les nœuds situés aux extrémités du réseau. Ceci est normal car l'algorithme de la multilateration n'obtient une bonne performance de localisation que lorsque les landmarks sont situés aux extrémités du réseau de façon à entourer l'ensemble des nœuds aveugles à localiser. Cette remarque est valable aussi pour les algorithmes basés sur MDS (MDS classique et MDS moderne) lorsqu'on utilise des landmarks pour avoir un référentiel absolu. L'algorithme DSAL quant à elle obtient de bonnes performances de localisation pour l'ensemble des nœuds individuels aveugles du réseau. L'erreur moyenne de localisation est inférieure à 0.72 m ou inférieur à 0.076 m en erreur moyenne normalisée pour tous les nœuds aveugles du réseau exceptés ceux dont les identifiants sont ID 247 et ID 198. En effet ces nœuds obtiennent une plus petite erreur moyenne de distance et une erreur moyenne normalisée de localisation respectivement ($NME = 0.063$ et $NME = 0.02$) avec MDS moderne et MDS classique. Toutefois il est à noter que le nœud dont l'identifiant est 247 est celui dont la position a été estimée avec la plus grande valeur de variance de bruit : $\sigma^2 = 14.09$ et la plus petite valeur de paramètre de perte de trajet $\eta = 0.66$.

Les résultats des simulations et d'expériences en environnements réels ont démontré une meilleure performance significative de localisation avec l'algorithme DSAL comparé à la multilateration, au MDS classique et au MDS moderne. En moyenne une performance de localisation de 0.77 m est observée sur une surface de 8 m par 5 m sous la présence de bruit. Cette performance est encore meilleure lorsqu'une initialisation avec un algorithme différent de la méthode aléatoire, exemple MLE, est considérée. DSAL est donc capable d'affiner et d'améliorer la performance de localisation de 85% comparée à ces algorithmes précités.

DSAL est également capable d'affiner la précision de localisation initialement estimée par B-MLE que nous avons proposé (cf. figures 6.28 et 6.29). Sur le testbed de Lincs, une erreur de localisation de 0.775 m et 1.3 m est obtenue en y appliquant DSAL sur les résultats de B-MLE respectivement de 1.39 m et 1.35 m pour la petite (i.e. 10 positions) et grande (i.e. 25 positions) base d'apprentissage. Sur le testbed Rammus Rennes, DSAL affine la précision de

localisation de 1.73 m et 1.64 m respectivement par rapport à B-MLE qui obtient une précision de localisation de 2.73 m et 2.20 m pour la petite (i.e. 10 positions) et grande (i.e. 25 positions) base d'apprentissage.

Nous remarquons ainsi sur l'ensemble des positions à localiser de nœuds aveugles sur le testbed Lincs que DSAL a affiné la précision de localisation de 79% et 55% par rapport à B-MLE respectivement pour la petite (i.e. 10 positions) et grande (i.e. 25 positions) base d'apprentissage. Cette précision va de 72% à 74% sur le testbed Rammus par rapport à B-MLE pour la petite (i.e. 10 positions) et grande (i.e. 25 positions) base d'apprentissage (cf. Tableau 6.6).

TABLE 6.6 – Résultats de la localisation de l'algorithme DSAL en utilisant un pas décroissant $((\gamma_t)_{t\geq 1}, \gamma_t = 1/\sqrt{t})$ et une initiation avec notre précédente proposition B-MLE.

Caractéristiques	**Testbed Lincs-petite base d'apprentissage**	**Testbed Lincs-grande base d'apprentissage**	**Testbed Rammus-petite base d'apprentissage**	**Testbed Rammus-grande base d'apprentissage**
Nombre de positions inconnues de nœuds aveugles	38	22	47	27
Nombre de positions de nœuds aveugles pour lesquels l'erreur de localisation n'est pas améliorée avec DSAL	8	10	13	7
Nombre de positions de nœuds aveugles pour lesquels l'erreur de localisation est significativement améliorée avec DSAL	30	12	34	20
Pourcentage de position dont l'erreur de localisation est nettement améliorée (plus petite) avec DSAL par rapport à B-MLE	79%	55%	72%	74%

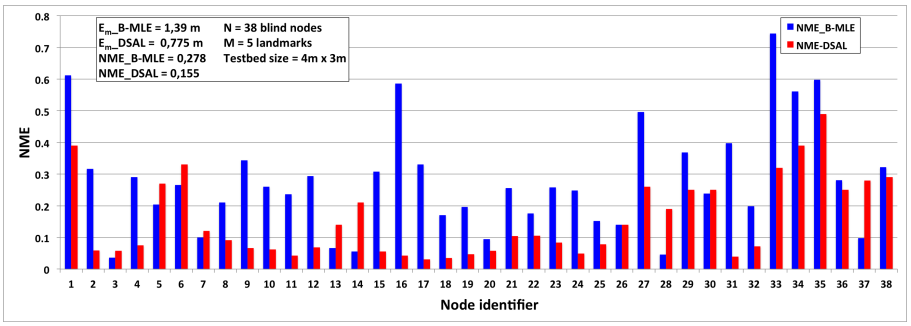
Conclusion

Nous avons réalisé des campagnes de mesures dans différents environnements internes afin de comprendre le comportement du RSSI en interne. Ces résultats de campagne de mesures sont disponibles dans mon site web[1].

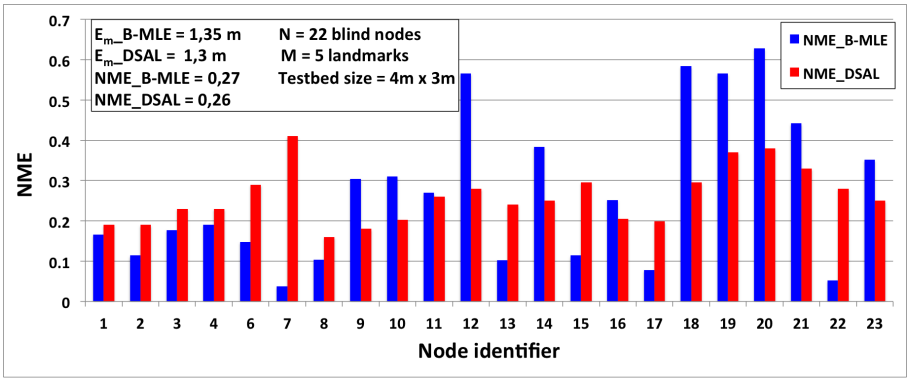
Nous avons ensuite proposé des algorithmes pour prendre en compte les incohérences observées du milieu interne, provenant de nos résultats des campagnes de mesures. Ces incohérences identifiées sont les effets induits par les multi-trajets et des valeurs aberrantes. Nous avons ensuite démontré que ces algorithmes sont adaptables à tout environnement interne sans calibration et sont capables d'améliorer la performance de localisation sous la présence de bruit, sans unité centrale de traitement et sans intervention humaine.

Grâce à ces filtrages statistiques appliqués sur les valeurs de RSSI mesurées et malgré quelques

[1] `http://nadieng.wordpress.com/`

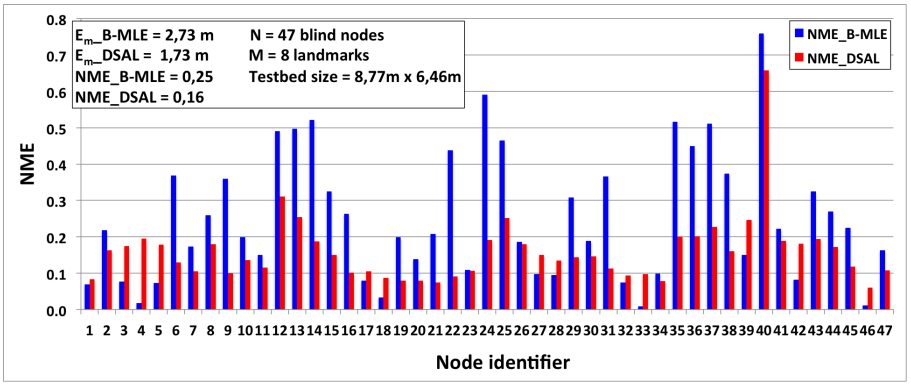


(a) Performance de localisation de l'algorithme DSAL avec un pas décroissant $((\gamma_t)_{t\geq 1},\ \gamma_t = 1/\sqrt{t})$ et une initialisation avec B-MLE. Résultats du testbed Lincs avec une petite base d'apprentissage de 10 positions.

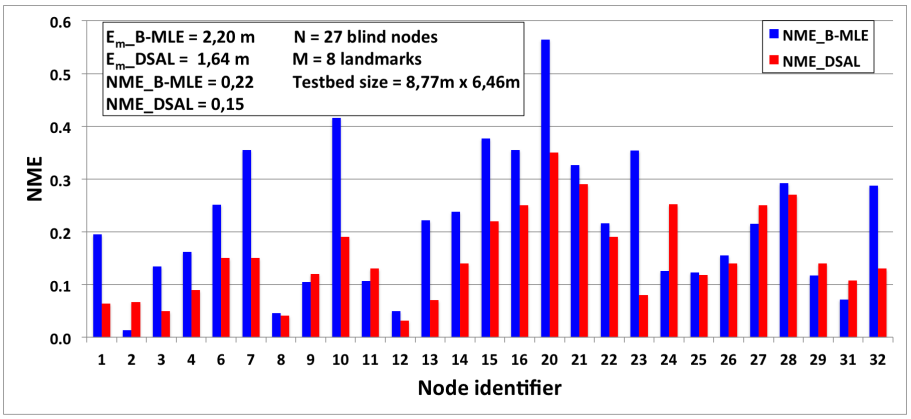


(b) Performance de localisation de l'algorithme DSAL avec un pas décroissant $((\gamma_t)_{t\geq 1},\ \gamma_t = 1/\sqrt{t})$ et une initialisation avec B-MLE. Résultats du testbed Lincs avec une base d'apprentissage plus grande, i.e. 25 positions.

FIGURE 6.28 – Performance de localisation de l'algorithme DSAL avec un pas décroissant $((\gamma_t)_{t\geq 1},\ \gamma_t = 1/\sqrt{t})$ et une initialisation avec B-MLE. Résultats du testbed Lincs.



(a) Performance de localisation de l'algorithme DSAL avec un pas décroissant ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) et une initialisation avec B-MLE. Résultats du testbed Rammus avec une base d'apprentissage petite de 10 positions.



(b) Performance de localisation de l'algorithme DSAL avec un pas décroissant ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) et une initialisation avec B-MLE. Résultats du testbed Rammus avec une base d'apprentissage plus grande, i.e. 25 positions.

FIGURE 6.29 – Performance de localisation de l'algorithme DSAL avec un pas décroissant ($(\gamma_t)_{t\geq 1}$, $\gamma_t = 1/\sqrt{t}$) et une initialisation avec B-MLE. Résultats du testbed Rammus.

erreurs, nous arrivons à avoir une précision de localisation suffisante pour le type d'application souhaité. Grâce à nos améliorations nous pouvons suivre une personne dans sa maison. De plus, nos algorithmes de localisation proposés peuvent également permettre de localiser des bébés dans une nursery, positionner une caméra avec précision devant une personne ou encore surveiller des œuvres d'art dans un musée.

Outre ces algorithmes de localisation à un saut proposé, nous avons également introduit un algorithme distribué pour affiner de façon efficace et significative la précision de localisation des nœuds d'un réseau de capteur sans fil.

Nos algorithmes proposés sont capables d'améliorer la performance de localisation en environnement interne. Les contributions de nos travaux sont résumées ci-dessous. Nous observons une amélioration de :

- 30% avec GMM-MLE par rapport à B-MLE
- 50% avec GMM-MLE par rapport à trilateration
- 40% avec B-MLE par rapport à MLE
- 20% avec B-MLE par rapport à trilateration
- en moyenne de 70% avec DSAL par rapport à B-MLE
- 85% avec DSAL par rapport à la Multilateration et MDS

Travaux futurs

Dans la mesure où nous avons proposé différents algorithmes pour prendre en compte différentes incohérences, une extension immédiate de notre travail est la localisation opportuniste.

Avec la localisation opportuniste il devrait être possible d'identifier en ligne le scénario d'usage pour une sélection intelligente entre les différentes méthodes de localisation disponibles. Il s'agit par exemple d'identifier certains critères à lier à la sélection des algorithmes : la taille du testbed, la présence de points d'accès sans fil. Nous croyons qu'une telle décision pourrait être basée sur la perception d'un seul nœud, provenant par exemple des propriétés statistiques de l'ensemble des données (i.e. « dataset »en anglais).

Une autre extension directement liée est d'utiliser cette localisation opportuniste pour optimiser la formation de groupe au sein d'un réseau de capteur sans fil. Par exemple grâce à la localisation opportuniste, on peut identifier plus facilement les nœuds proches ou les nœuds qui bougent ensembles.

Différents groupes peuvent ensuite échanger et combiner des informations (par exemple, les paramètres du modèle de propagation, la distance entre les nœuds) pour améliorer encore leur précision suivant le réseau de capteur global. Une telle technique permet de traiter des données localement car on est plus proche de la source. Une gestion du mécanisme d'adressage des nœuds et des préfixes d'adressage IP peut aussi être rendue plus souple pour le passage à l'échelle du système d'adressage.

Ainsi, l'allocation des préfixes globaux d'adressage IP peut être améliorée. Les préfixes d'adresses IP peuvent être alloués à des clusters et le « cluster head »(chef de groupe) va s'occuper d'attribuer ces adresses IP aux nœuds. De plus, avec une localisation précise, l'identifiant du « cluster

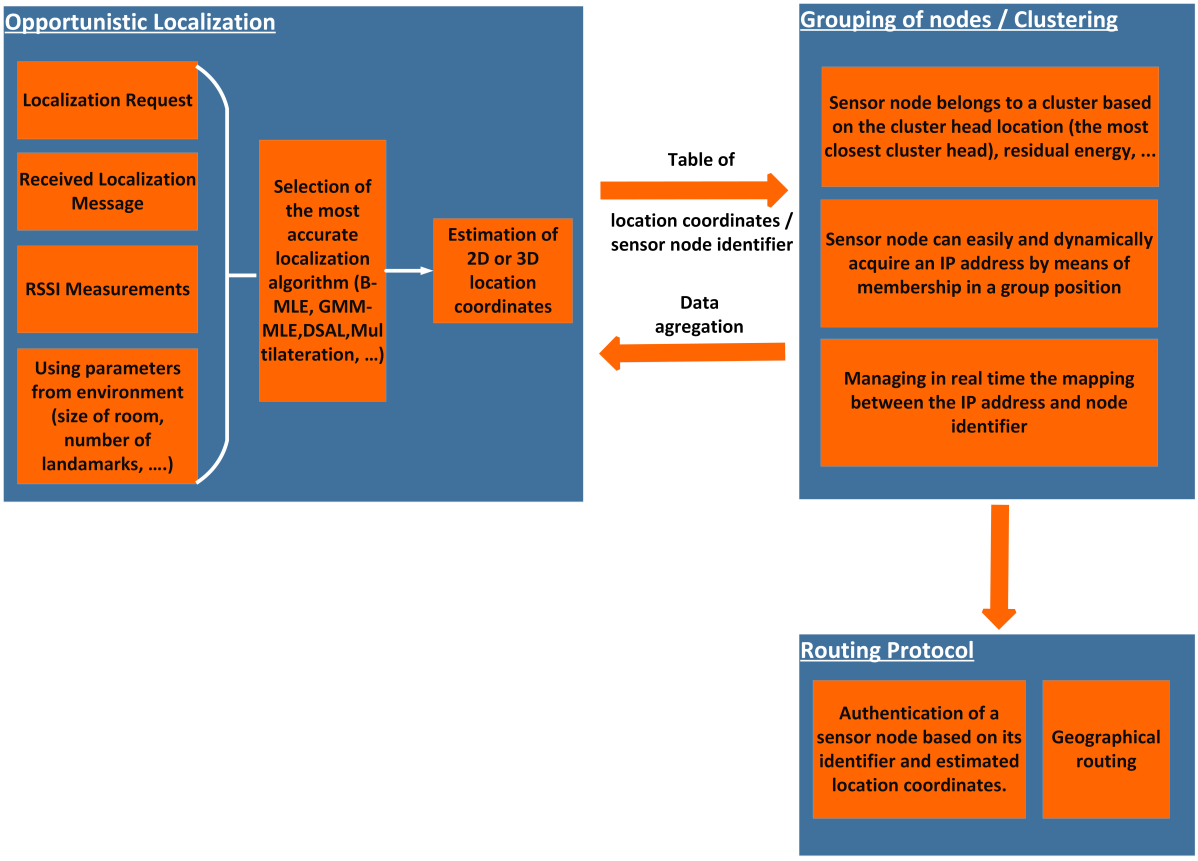


FIGURE 6.30 – Architecture proposée de regroupement de nœuds capteurs ou « clustering »basé sur la localisation pour améliorer le routage dans un réseau de capteurs sans fil multi-sauts

head »peut être associé à la position localisée pour former ensemble une information stable, partagée par des nœuds proches ou des nœuds qui bougent ensemble. Localiser des nœuds capteurs qui sont proches permet d'avoir des groupes/clusters et permet ainsi le routage des données entre clusters. Pour aller à une destination, la route peut être identifiée de façon macroscopique, en envoyant les informations au groupe (« cluster head »ou un des nœuds du groupe). Ensuite n'importe quel nœud du groupe ayant reçu la donnée peut le router au groupe suivant. Ce qui contribue à limiter l'envoi de paquets de routage dans l'ensemble du réseau et permet la scalabilité.

Dans cette perspective, l'auto-organisation d'un réseau par le biais du regroupement des nœuds associé à la localisation opportuniste permet d'améliorer le routage multi-saut au sein d'un réseau de capteur (cf. figure 6.30). De plus, la position ou le chemin de localisation du nœud mobile aveugle, caractéristiques unique de ce dernier, peuvent être exploité par un protocole de routage à des fins d'authentification avant l'envoi des données.

En outre, puisque les capteurs communiquent des données associées à une faible distance dans un environnement de groupe/cluster, l'énergie consommée par le réseau sera inférieure à celle consommée sans formation groupe/cluster.

Appendix A

Comparison of RSSI-based indoor localization mechanisms

A.1 Comparison of RSSI-based indoor localization mechanisms

According to Table A.1, non-distributed algorithms do not mean they use centralized approach. Instead, most one-hop based algorithms do not use a base station (centralized approach), as the localization can be performed by a blind node itself. In addition, almost all algorithms use the log-normal shadowing model as wireless propagation model. If this is not the case, we mention the used propagation model in the field of the accuracy.

Table A.1: Comparison of RSSI-based indoor localization mechanisms

Reference	**Control of node placement**	**Experiment-based**	**Accuracy in meters**	**Type of algorithm**	**Used approach**
[120]	Each node is separated from its neighbor by an angle of 120°	yes	2 m within 20 m	1 hop non-distributed localization with 3 landmarks	Gaussian distribution function
[99]	None	yes	2.13 m for variable office and 1.13 m for non obstructed office. (size of 18 m by 18 m)	Centralized approach with 4 landmark, a base station and 10 blind nodes positions	Estimation of the path loss exponent according to each environment: Path loss exponent estimation algorithm based on polygon method

Continued on next page

Table A.1 – continued from previous page

Reference	Control of node placement	Experiment-based	Accuracy in meters	Type of algorithm	Used approach
[123]	Adaptive testbed for distance measurement, control of node antenna orientation	yes	0.50 cm of accuracy in an area of 3.5 m by 4.5 m	1 hop non-distributed localization with 7 landmarks and 6 blind nodes positions	RSSI-based localization in WSN which controls and exploits indoor influenced factors to improve measurements results (adjustment of transmission power, frequency, antenna orientation, quality of the reference measurements). Distance estimation between nodes from measured RSSI with Artificial neural networks and trilateration mechanism
[164]	None	No	Need at least 12 landmarks to obtain localization error less than 150 cm in specific single room scenario of size 7 m by 4.95 m and a height of 3.12 m	One-hop localization with MLE using Montecarlo simulation	One-hop localization with MLE using Montecarlo simulation. For each grid (mobile node position) computing the likelihood that the mobile node is there
[6]	No	yes	Better positioning accuracy of 3 m with min-max in a testbed of size 12 m by 19.5 m, node density of 0.013 nodes squared meter, minimum is 0.08 nodes per squared meter. Instead, using two antennas (Mimo) spaced 10cm, on a receiver and read RSSI alternatively from both, increases positioning accuracy by 20 %	1 hop non-distributed localization with 3 landmarks and 3 blind nodes positions, repeated 25 times	Comparison of three ranging algorithms: Geometric consideration (Min-max, trilateration) and statistics (MLE). RSSI values cleaned from outliers

Continued on next page

Table A.1 – continued from previous page

Reference	Control of node placement	Experiment-based	Accuracy in meters	Type of algorithm	Used approach
[5]	No	Yes	Better positioning accuracy of 2.5-3 m with MLE in a testbed of size 10 m by 10 m with 7 landmarks. Instead MLE require more much computational coast then min-max. Performance of multilateration and ROCRSSI are comparable and worst than the other schemes	1 hop non-distributed localization with 48 landmarks at height 75 cm from ceiling and 12 landmarks (2nd testbed composed of WSN) 3 blind nodes positions	Channel characterization and error correlation techniques are exploited performance comparison of Mix-max, ROCRSSI, trilateration and MLE in indoor WSN
[102]	No	Yes	Conference room (area: 7.08 m by 10.60 m), density of 0.27 nodes per squared meter leads to positioning error of 1.5-2 m	Centralized approach, position estimation on a sink node with 20 landmarks and 6 blind nodes positions	MLE based indoor localization system with the use of the RSSI threshold (based on proposed linear equation between RSSI and distance) for data filtering and an effective data collection methodology for accuracy improvement: a user decides the number of data to collect based on prior knowledge (proposed density function) and send it to sensor nodes by flooding from the sink node
[7]	No	Yes	Better results obtained with maximum transit power (0dBm), at Strasbourg platform : 2 m of localization error using RSSI-based distance. Spring method do not require here to increase the landmarks as localization accuracy decrease with more that 50 landmarks	Spring-based 1 hop collaborative non-centralized algorithm. Evaluation in 3 large scale sensor network testbeds of 256 nodes each	RSSI-based spring algorithm in large scale sensor network testbed (i.e. SensLAB)

Table A.2: Comparison of RSSI-based localization algorithms accuracy

References	**Authors**	**Title (year)**	**Number of nodes (N)**	**Number of Landmarks (M)**	**Accuracy (Em) in meters**	**Dimensions (P,Q) in meters**	**NME as $\frac{Em}{\sqrt{(P^2+Q^2)}}$**
[29]	Dragos Niculescu *et al.*	Ad hoc positioning system (APS) (2001)	100	≥ 80	3	10×10	0.2121
[5]	Giovanni Zanca *et al.*	Experimental comparison of RSSI-based localization algorithms for indoor wireless sensor networks (2008)	48	25	2.2	10×10	0.1556
[92]	Jun Zheng *et al.*	Range-free localization in wireless sensor networks with neural network ensembles (2012)	250	75	2	50×50	0.0283
[86]	Chong Liu *et al.*	Range free sensor localization with ring overlapping based on comparison of received signal strength indicator (2007)	-	-	5.18	40×30	0.1036
[7]	Karel Heurtefeux *et al.*	Is RSSI a good choice for localization in wireless sensor network? (2012)	206	50	2.2	11×9	0.1548
[123]	Abdalkarim Awad *et al.*	Adaptive distance estimation and localization in WSN using RSSI measures (2007)	7	6	0.5	3.5×4.5	0.0877
[203]	Mehdi Essoloh *et al.*	Localisation distribuée dans les réseaux de capteurs sans fil par résolution d'un problème quadratique (2007)	72	8	5.12	90×90	0.0402
[125]	Jose A. Costa *et al.*	Distributed Weighted-Multidimensional Scaling for Node Localization in Sensor Networks (2005)	40	4	2.269	14×14	0.1146
[159]	Neal Patwari *et al.*	Relative Location Estimation in Wireless Sensor Networks (2003)	12 40	4 4	1.02 2.18	9×9 14×13	0.0801 0.1141

Appendix B

Short survey on Spring algorithms

Contents

B.1 Short survey on Spring algorithms

Using the concept of graph, the spring algorithm is defined as a network of nodes connected by spring. In the spring localization mechanism, forces (based on RSSI measurements) are applied to all nodes in the network. Applied forces attract or repulse nodes each other until the system reaches its equilibrium state. In other words, the spring algorithm minimizes springs extension or compression and thus uses a local minima: initial sensor location are randomly assigned by the designer. The wireless sensor nodes are then modeled as particles connected by a network of springs: The length of the spring is the distance between nodes. Distances between each sensor node and the neighboring landmarks are estimated from RSSI measurements to the landmarks. Additionally, distance between blind nodes are measured [204]. The estimated distances between sensor nodes are then iteratively used to refine the position of each nodes within the network of spring. Note that measured RSSI values used to estimate the distance are averaged within 3 hours. Finally, the position error difference in meters between the real position and the estimated position is computed for each node and defined as the localization accuracy [7].

The proposed RSSI-based spring relaxation algorithm has been recently implemented and tested in SensLAB large scale wireless sensor network platform with 700 nodes deployed in 3 different sites with TI CC11101 and CC2420 802.15.4 RF nodes [7]. Using the maximum transmit power (i.e. 0 dBm) which is the best condition for linear RSSI according to the authors experimental results, an average localization error of 3 meters is obtained on the first SensLAB platform named Grenoble platform. Authors link this result to a larger deployment area that leads to a worst initial random location of the RSSI-based spring algorithm. However, a comparable performance is observed on the second SensLAB platform named Lille. Furthermore, on the third SensLAB platform named Strasbourg and of size 11 m by 9 m, a localization error of about 2.2 m is obtained in 2D dimension while less than several centimeters is reported when using euclidean distance instead of RSSI-based distance estimation [7]. The used number of landmarks for each testing platform is not given by the authors. Instead, authors state that the increase of the number of landmarks is not needful when the total number of $N = 50$ landmarks is reached. In fact, their results demonstrated that the usage of more than 50 landmarks degrades the average localization accuracy. Authors qualify such deterioration to come from poor distance estimation from RSSI readings. Instead, a prior work from [94] have demonstrated with simulation that increasing the number of sensors involves more sensors'locations to be estimated with the spring relaxation localization technique; because the network connectivity (number of neighbors in this case) will increase. In addition, when evaluated the spring algorithm in simulation using two-hops network of size 350 m by 350 m square map, the average localization error diminishes from

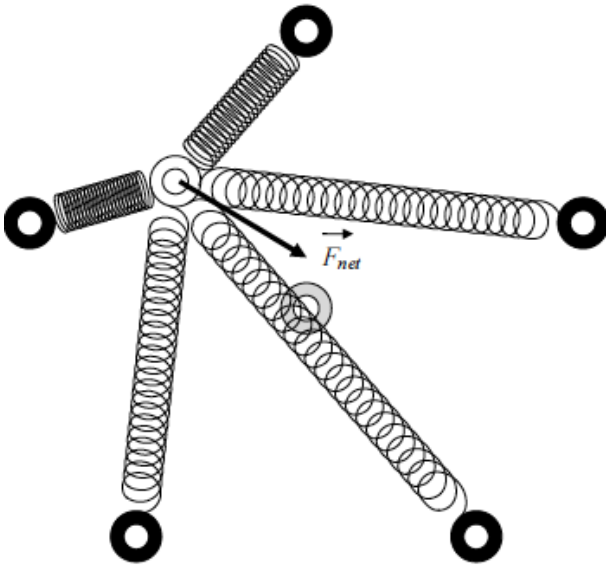


Figure B.1: Overview of a wireless network connected by springs using five landmarks (represented by black circle) and a blind node [94]

42.4 m to 22.63 m when increasing the number of sensor nodes: i.e. number of neighbors from 100 (connectivity of 4.13 neighbors for a node and $unsolvedlocation = 13\%$) to 600 (connectivity of 8.82 neighbors for a node and $unsolvedlocation = 0.33\%$). The used number of landmarks in simulation is about 64.24% of sensors with one-hop to some landmarks, and about 35.76% sensors with two-hops cases. Undoubtedly the result of spring layout degrades when the number of nodes increases [126]. In [204] an accuracy of 3 m is obtained using the spring method in an indoor testbed of size 34.6 m by 18.3 m at Berkeley. This accuracy of 3 m is obtained with Micaz nodes composed of three landmarks and 28 receiver nodes.

This better accuracy observed in the works presented in [204] and compared to the one in [7] that almost use the same distributed localization mechanism can be explained by the used method to infer distance from measured RSSI values. In [204], measured RSSI are converted to distance using a graph available online from [205] performed on the same CC1000 radio chip used by mica2dot [204]. Based on the authors, *mica2dot* has a slower performance than *Telos* node and can not perform time of flight measurement.

Instead, in [7], the path loss model to map measured RSSI values to distances is Friis radio propagation. Or, Friis path loss, presented earlier in Chapter 2, is not designed for the use in indoor environment. From the authors, the spring relaxation algorithm has two drawbacks that can affect its effectiveness: the first one is the used local minima needed by the algorithm to work. Actually, a used local minima found by the algorithm during its execution can be significantly worst than the global minima. The second drawback that affects the Spring relaxation algorithm is its running time complexity equivalent to $\mathcal{O}(n^3)$ and the number of iteration which is $\mathcal{O}(n)$, n corresponds to the number of the sensor nodes in the wireless sensor network. From the authors conclusion, Range-based RSSI method is not a good candidate for distance estimation in wireless sensor network. They suggest to investigate connectivity based range-free algorithms as they can give useful metrics instead of RSSI-distance estimation. However, as we demonstrated in Chapter 2 in Section 2.2.3, most range-free localization algorithms provide coarse location accuracy which is not sufficient when it comes to to track a person in an indoor environment.

Appendix C

Selection of a propagation model

Contents

C.1 Selected propagation model

Additional path loss models have been investigated and presented in several works. Such a work have concentrated on modifying the log-distance model by adding some encountered situation loss factors from different indoor environments. In fact, an adaptive two-ray model has been utilized in [175] (cf. the equations C.1, C.2) when considering the path loss as the (logarithmic) difference between an emitted power at 0 dBm and a receiving power while reading the received power (RSSI) in dBm.

For the distance $d \leq 8$

$$\mathrm{PL}(d) = 40.2 + 20\log_{10}(d), \tag{C.1}$$

And for the distance $d \geq 8$

$$\mathrm{PL}(d) = 58.5 + 33\log_{10}(\frac{d}{8}), \tag{C.2}$$

The authors use the above two ray model based on the assumption that many works like in [54] have added an error correction to the equations C.1, C.2, that is WAF for the wall attenuation factors and FAF for the floor attenuation factors. However such parameters are then minimized in their investigation. This is the case in [54] where the floor attenuation factor is minimized while considering only the walls attenuation by assuming the localization system to work in one floor. However, many walls in modern building are plywood, thus traversing them does not attenuate a lot RSSI readings. However, the features, the objects and their layout within the environment considerably affect the RSSI measurements. Such effects are in most case modeled by the shadowing within the log-normal shadowing model. Compared to other propagation model, this latter model is the best candidate for indoor localization when taken into account the trade off between complexity and feasibility. The log-normal shadowing model which uses fewer parameters than the modification of the log-distance model with additional attenuation factors (partition losses between floors and walls), is claimed to be more accurate than the two-ray ground model and can be easily programmed on the current sensor nodes.

Appendix D

Trilateration

Contents

D.1 Trilateration *Lodachai*

Let denote by k a positive constant for the distance ratio $R_{i,j}$ between two landmarks and a mobile node. Let also use two points PT_i and PT_j. We will study all point PT with the following property: $d(PT, PT_i) = kd(PT, PT_j)$.

1. If $k = 1$ $d(PT, PT_i) = d(PT, PT_j)$

   In the orthogonal reference such as PT_i with coordinate $(0, 0)$ and PT_j with coordinate $(x_j, 0)$ and $x_j > 0$, the system of equations is:

   $$x^2 + y^2 = (x - x_j)^2 + y^2, \qquad x^2 = x^2 - 2xx_j + x_j^2, \qquad x = \frac{x_j^2}{2x_j}, \qquad x = \frac{x_j}{2} \tag{D.1}$$

   We show that the solution is the median of the segment $[PT_i, PT_j]$

2. If $k < 1$

   $d(PT, PT_i) = kd(PT, PT_j)$ Let $k\prime = \frac{1}{k}$ and $PT_i\prime = PT_j$ and $PT_j\prime = P_i$.

   Solving the $d(PT, PT_i) = kd(PT, PT_j)$ amounts to solve $d(PT, PT_j\prime) = \frac{d(PT, PT_i\prime)}{k\prime}$ or

   $d(PT, PT_i\prime) = k\prime d(PT, PT_j\prime)$ with $k\prime > 1$

   It is the same to solve the case $k > 1$ by reversing the points PT_i, PT_j

3. If $k > 1$

   $d(PT, PT_i) = kd(PT, PT_j)$

   In the orthogonal reference with both points with coordinates $PT_i(0, 0)$ and $PT_j(x_j, 0)$ with $x_j > 0$, the systems of equations is expressed as:

   $$x^2 + y^2 = \mathcal{K}(x - x_j)^2 + \mathcal{K}y^2 \quad with \quad \mathcal{K} = k^2, x^2 + y^2 = \mathcal{K}x^2 - 2\mathcal{K}x_jx + \mathcal{K}x_j^2 + \mathcal{K}y^2 \tag{D.2}$$

   $(1 - \mathcal{K})x^2 + (1 - \mathcal{K})y^2 + 2\mathcal{K}x_jx - \mathcal{K}x_j^2 = 0$

is of the form of

$\hat{\mathrm{E}}Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$

wtih $A = 1 - \mathcal{K}$, $B = 0$, $C = 1 - \mathcal{K}$, $D = 2\mathcal{K}x_j$, $E = 0$, $F = -\mathcal{K}x_j^2$.

The solution, if it exists, is part of the conical.

We will now study the function f in $\mathbb{R} \to \mathbb{R}$ which associates $(1 - \mathcal{K})x^2 + 2\mathcal{K}x_j x - \mathcal{K}x_j^2$ to x with $\mathcal{K} > 1$.

In $+\infty$ and $-\infty$, the function f tends to $-\infty$ for $1 - \mathcal{K} < 0$

Let solve the function $f(x) = 0$, that is $(1 - \mathcal{K})x^2 + 2\mathcal{K}x_j x - \mathcal{K}x_j^2 = 0$.

$$\Delta = 4\mathcal{K}^2 x_j^2 - 4(1 - \mathcal{K})(-\mathcal{K}x_j^2)\Delta = 4\mathcal{K}^2 x_j^2 + 4\mathcal{K}x_j^2 - 4\mathcal{K}^2 x_j^2\Delta = 4\mathcal{K}x_j^2 \tag{D.3}$$

$\Delta > 0$ because $\mathcal{K} > 0$ and $x_j > 0$

$r_i = \frac{-2\mathcal{K}x_j - \sqrt{(4\mathcal{K}x22)}}{2(1-\mathcal{K})}$

$r_j = \frac{-2\mathcal{K}x_j + \sqrt{(4\mathcal{K}x22)}}{2(1-\mathcal{K})}$

$r_i = \frac{-\mathcal{K}x_j - x_j\sqrt{(\mathcal{K})}}{(1-\mathcal{K})}$

$r_j = \frac{-\mathcal{K}x_j + x_j\sqrt{(\mathcal{K})}}{(1-\mathcal{K})}$

$r_i = \frac{(-\sqrt{\mathcal{K}} - \mathcal{K})x_j}{1-\mathcal{K}}$

$r_j = \frac{(\sqrt{\mathcal{K}} - \mathcal{K})x_j}{1-\mathcal{K}}$

if $x = r_i$ or $x = r_j$ then $f(x) = 0$

As $1 - \mathcal{K} < 0$ if $x < r_j < r_i$, $x - r_i > 0$, $x - r_j > 0$ and $(1 - K) < 0$ we have $(1 - \mathcal{K})(x - r_i) < 0$ and

$(1 - \mathcal{K})(x - r_i)(x - r_j) < 0$ then $f(x) < 0$

If $r_j < x < r_i$, $x - r_i > 0$, $x - r_j < 0$ and

$(1 - \mathcal{K}) < 0$ we have $(1 - \mathcal{K})(x - r_i) < 0$ and $(1 - \mathcal{K})(x - r_i)(x - r_j) > 0$ then $f(x) > 0$

if $r_j < r_i < x$, $x - r_i < 0$, $x - r_j < 0$ and $(1 - \mathcal{K}) < 0$ we have $(1 - \mathcal{K})(x - r_i) > 0$ and $(1 - \mathcal{K})(x - r_i)(x - r_j) < 0$ then $f(x) < 0$

Note that if $x = 0$, (in (PT_i)), $f(0) = -\mathcal{K}x_j^2 < 0$

if $x = x_j$, (in (PT_j)), $f(x_j) = x_j^2 > 0$

Let solve the derivative $f\prime$ of f $f\prime(x) = 2(1-\mathcal{K})x + 2\mathcal{K}x_j$ $f\prime(x) = 2((1-\mathcal{K})x + \mathcal{K}x_j)$ $f\prime(x) = 0 -> (1-\mathcal{K})x + \mathcal{K}x_j = 0$ $x = \frac{\mathcal{K}x_j}{(\mathcal{K}-1)}$

therefore the function f reaches its maximum when $x = \frac{\mathcal{K}x_j}{(\mathcal{K}-1)}$ Let us now solve the function g, $y \mapsto (\mathcal{K}-1)y^2$ with $\mathcal{K} > 1$

In $+\infty$ and $-\infty$, the function g tends to $+\infty$ for $(\mathcal{K}-1) > 1$

$g(y) = 0$ admits one and only one solution: $y = 0$ For all $y \neq 0$, $g(y) > 0$ for $\mathcal{K} - 1 > 0$ and $y^2 > 0$

Let us study the set (E) of couples (x, y) such as $g(y) = f(x)$

We know that for all y, $g(y) \geq 0$ and that $f(x)$ is positive if and only if $r_j \leq x \leq r_i$

The set of pairs (x, y) that satisfy $f(x) = g(y)$ is the set defined by:

$$\left((r_j \leq x \leq r_i), y = \pm\sqrt{\frac{(1-\mathcal{K})x^2 + 2\mathcal{K}x_j x - \mathcal{K}x_j^2}{(1-\mathcal{K})}}\right)$$

As the function f reaches its maximum at $(\frac{\mathcal{K}x_j}{(\mathcal{K}-1)})$, all (x, y) pairs are included in the rectangle defined by:

$$\left((r_j \leq x \leq r_i), -\sqrt{\frac{f_max x}{(1-\mathcal{K})}} \leq y \leq \sqrt{\frac{f_max x}{(1-\mathcal{K}}}\right)$$

with f_{max} being the maximum of the function f.

We are going to show that the set E is the solution of $d(PT, PT_i) = \mathcal{K}d(PT, PT_j)$

The equation $(1-\mathcal{K})x^2 + (1-\mathcal{K})y^2 + 2\mathcal{K}x_j x - \mathcal{K}x_j^2 = 0$ found above can be written as: $(1-\mathcal{K})x^2 + 2\mathcal{K}x_j x - \mathcal{K}x_j^2 = (\mathcal{K}-1)y^2$ which corresponds to the expression $f(x) = g(y)$ The set E then corresponds to the solution of $d(PT, PT_i) = \mathcal{K}d(PT, PT_j)$ with $\mathcal{K} = k^2$

Let study the nature of the set E:

We have seen that the set E is a conical bounded in the plane by a rectangle. The only conical meeting this requirement is an ellipse [206].

The set of points in $d(PT, PT_i) = \mathcal{K}d(PT, PT_j)$ with $\mathcal{K} > 1$ is an ellipse. This ellipse surrounds the point PT_j but not the point PT_i; (PT_j being inside and PT_i being outside).

Studying a little further:

$(1-\mathcal{K})x^2 + (1-\mathcal{K})y^2 + 2\mathcal{K}x_j x - \mathcal{K}x_j^2 = 0$

$x^2 + y^2 + \frac{2\mathcal{K}x_j x}{(1-\mathcal{K})} - \frac{\mathcal{K}x_j^2}{(1-\mathcal{K})} = 0$

$x^2 + \frac{2\mathcal{K}x_j x}{(1-\mathcal{K})} + \frac{\mathcal{K}^2 x_j^2}{(1-\mathcal{K})^2} - \frac{\mathcal{K}^2 x_j^2}{(1-\mathcal{K})^2} - \frac{\mathcal{K}x_j^2}{(1-\mathcal{K})} + y^2 = 0$

$\frac{x+\mathcal{K}x_j}{(1-\mathcal{K})^2} + (y-0)^2 = \frac{\mathcal{K}^2 x_j^2}{(1-\mathcal{K})^2} + \frac{\mathcal{K}x_j^2}{(1-\mathcal{K})}$

$\frac{x-\mathcal{K}x_j}{(\mathcal{K}-1)^2} + (y-0)^2 = \frac{(\mathcal{K}^2x_j^2+\mathcal{K}x_j^2(1-K))}{(1-\mathcal{K})^2}$

$\frac{x-\mathcal{K}x_j}{(\mathcal{K}-1)^2} + (y-0)^2 = \frac{(\mathcal{K}^2x_j^2+\mathcal{K}x_j^2-\mathcal{K}^2x_j^2)}{(1-\mathcal{K})^2}$

$\frac{x-\mathcal{K}x_j}{(\mathcal{K}-1)^2} + (y-0)^2 = \frac{\mathcal{K}x_j^2}{(\mathcal{K}-1)^2}$

This is the equation of a circle with center $(\frac{k^2x_j}{(k^2-1)}, 0)$ and radius $\frac{k*d(PT_i,PT_j)}{\sqrt{(k^2-1)^2}}$.

The radius of the circle is simply estimated by calculating this value: $\frac{k*d(PT_i,PT_j)}{\sqrt{(k^2-1)^2}}$.

For the center of the circle C, we have: the vector $P_iC = \frac{k^2}{(k^2-1)}$, the vector PT_iPT_j

if PT_i coordinate is (X_i, Y_i) and $PT_j(X_j, Y_j)$ then

$X_c - X_i = (\frac{k^2}{(k^2-1)})(X_j - X_i)$

$Y_c - Y_i = (\frac{k^2}{(k^2-1)})(Y_j - Y_i)$

$X_c = \frac{(k^2X_j-X_i)}{(k^2-1)}$

$Y_c = \frac{(k^2Y_j-Y_i)}{(k^2-1)}$

The solution of the conical is:

Let us denote two points $PT_1(X_1, Y_1)$ and $P_2(X_2, Y_2)$; the set of points PT such as $d(PT, PT_1) = kd(PT, PT_2)$ is a circle of center C $(\frac{(k^2X_2-X_1)}{(k^2-1)}, \frac{(k^2Y_2-Y_1)}{(k^2-1)})$ and radius $(\frac{k*d(P_1,P_2)}{\sqrt{(k^2-1)^2}})$.

Appendix E

Localization metrics estimation

Contents

E.1 Localization metrics estimation

RSSI is an available measurement on embedded devices. Angle of arrival metric can be estimated using the Fourier transform by separating sinusoids resulting from plane waves. The Fourier transform is a mathematical tool for signal processing used to analyze spectra. In other words, the Fourier transform is an operation that transforms an integrable function on $\mathbb{R}$ to another function while describing the frequency spectrum of this last. If f is an integrable function on $\mathbb{R}$ its Fourier transform is the function $F(u)$ given by the following formula in equation E.1:

$$F(u) = \int_{-\infty}^{\infty} f(x) \exp -i2\pi ux \, \mathrm{d}x \tag{E.1}$$

Thus a temporal/time-related function $f(t)$ can be associated through a Fourier transform to a complex-valued function $F(u)$ called spectrum of $f(t)$. This spectrum contains the weight of each of the frequencies u in the original function [207].

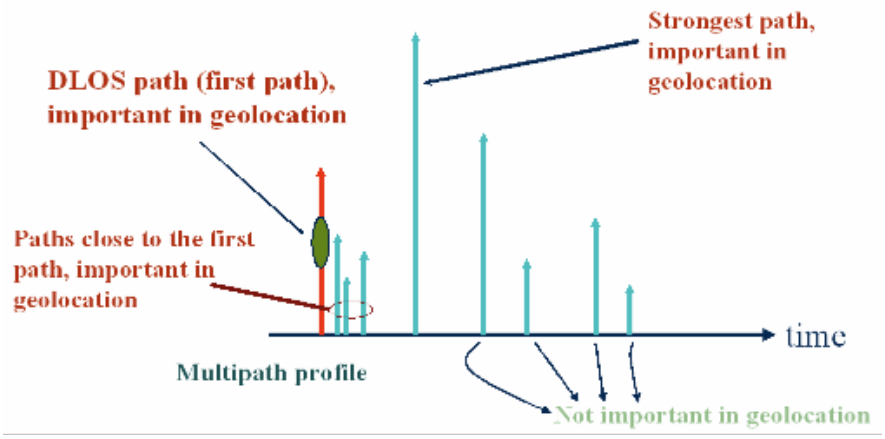
Similarly the inverse Fourier transform is defined in equation E.2 expressed as:

$$f(x) = \int_{-\infty}^{\infty} F(u) \exp +i2\pi ux \, \mathrm{d}u \tag{E.2}$$

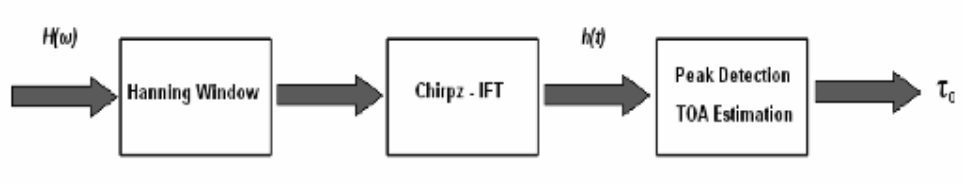
The estimation of a direction of angles with Fourier transform follows this procedure: a frequency representation is first performed for a given wave. Then there is a display of an energy peak for the direction that the source is. In case of a superposition of several waves, there are peaks for each Direction of Angle (i.e. DoA) [208]. In addition, to be able to distinguish with Fourier transform two sinusoids resulting from two planar waves with angles of arrival $\theta 1$ and $\theta 2$, the window between the two sinusoids should be necessary wide enough [208] and the requirement is given in the equation(E.3).

$$\|\sin\theta_1 - \sin\theta_2\| \geq \frac{\lambda}{2\pi d} \tag{E.3}$$

Instead, the estimation of time of arrival (ToA) can be done with inverse Fourier transform (cf. equation E.2) by first determining the correlation of received signals on a receiver; and then identifying the first path (based on precise clock synchronization) [106], [104]. Note that the



(a) An example of ToA profile in multipath situation



(b) Overview of Inverse Fourier Transform algorithm for ToA estimation

Figure E.1: An example of Time of Arrival (ToA) estimation [106]

correlation between two objects helps to measure their mutual dependence. The autocorrelation of a signal thus measures the internal dependencies of the signal [209, 210]. Estimating the time of arrival from the channel profile requires several steps as presented in Figures E.1(a) and (b):

Because of the presence of noise and multipath induced by reflections from the obstacles, the direct path does not always correspond to the strongest signal. Therefore the estimation of ToA from the channel profile requires several steps and an example is shown in Figure E.1. In fact the obtained transfer function is first passed through a *Hanning* window to reduce the secondary lobes of the impulse response while preventing losses and false peaks. Then an inverse Fourier transform is performed on the filtered impulse response to move to the time domain. The ToA is then estimated for each peak corresponding to the Direct Line of Sight (DLoS). Finally, the peak detection algorithm will select the first peak corresponding to the current ToA. An impulse response of a discrete or sampled system is the system output when the input is an impulse. Also called the temporal response, an impulse is mathematically modeled by a Dirac function [211] presented in equations E.1 and E.4. The impulse response is then the output of a system in response to an impulse. A filter is a linear system whose input-output behavior is characterized by the coefficients of its impulse response.

$$\delta(t) = \begin{cases} 0 & \text{si } t \neq 0 \\ +\infty & \text{si } t = 0 \end{cases}$$

and

$$\int_{-\infty}^{+\infty} \delta(t)\,\mathrm{d}t = 1 \tag{E.4}$$

Fourier analysis consists of representing the signal in a transformed domain: from time to frequency domain. It has been used in the fields of radar and sonar to detect sources, but it does not provide sufficient accuracy [208]. As Fourier cannot detect two closed sinusoids, this leads to non efficient technique on detecting arrival angles while showing the necessity to move toward more precise techniques. To achieve better accuracy and overcome the limits of Fourier transforms, some improvements have been made and they are based on (downstream) signal processing of antennas: these are high resolution methods (HR). ToA and AoA (i.e. Angle of Arrival) can also be estimated using such high resolution methods which are parametric estimation methods applied to antennas treatment. The following section is not intended to present complete HR methods. We first want to first notify the reader with the concepts on which they are based. Second we want to inform about the existence of these approaches being robust to estimate the time of arrival or the angle of arrival of the received signal, two alternatives metric to the measurements of signal attenuation.

High resolution methods estimate the parameters of a signal which usually consists of a sum of sinusoids modulated exponentially and disturbed by additive noise. The total signal received at the antenna array is the sum of signals received at different sensor antennas. With high resolution methods (for a plane wave approximation), it is possible to distinguish two closed sinusoids which is not able to perform the Fourier transform [208]. The high resolution methods can be used in practice with a shorter window than that usually applied in the Fourier analysis.

High resolution methods have many variants [208]:

- The *Prony* method or the linear prediction approach is the first approach from high resolution methods. An exponential sum is estimated by linear prediction approach. The output signal from the sensors or the antenna array is first decomposed into a sum of exponential. Then the obtained exponential samples satisfy the relationship that will allow to deduce a system of multiple equations with multiple unknowns. The resolution of this equation system in the absence of noise gives the Direction of Arrival angle estimation (DoAS). However the *Prony* method is very sensitive to noise.

- The CAPON method or the maximum likelihood performs an estimate without bias and minimum variance. From several samples of the received signals, an estimation of the amplitude of sources is performed depending on the direction of the angle.

- The quadratic minimization method minimizes a criterion for the maximum likelihood with the theoretical results. When applied to a sensor network, the autocorrelation matrix is first estimated; then the value of the angle by which the standard Least Squares recursive algorithm (i.e. MCR) is minimal. The corresponding angle is given by the MCR.

- Subspace-based methods are those based on the calculation of the auto-correlation matrix of the received signals on the different sensors. The subspace-based methods decompose the observation space of the sensor array (i.e. wherein the signal evolves) into two subspaces: a signal subspace and a noise subspace. This decomposition is made by research of eigenvalues as presented in Figure E.2. Such decomposition shows that the smaller area corresponds to the noise. The signal space is generated by the sinus curves and noise subspace is its orthogonal complement. This decomposition is also possible through a singular value decomposition (i.e. SVD) of the data matrix or eigenvalue decomposition (i.e. EVD) of the covariance matrix of the signal.

  Eigenvectors and eigenvalues are part of linear algebra and aim to decompose efficiently the space in direct sum of stable subspaces [212, 213]. The eigenvectors are the vectors

whose direction is unchanged by the linear application of a vector space into itself. A singular value decomposition is a factorization of a rectangular real or complex matrix. The singular value decomposition is general and can be applied to any rectangular matrix. Instead the eigenvalue decomposition applies only to certain classes of square matrices. Note that these decompositions exist in practice with iterative algorithms.

Subspace methods are often used when localizing emitters nodes by a network of sensors as shown in Figure E.3: estimation of the angle of arrival of the received signal.

For example, consider an antenna array of M sensors nodes and N emitters nodes with $N \leq M$. The signal without noise from a receiver entering the linear antenna array at time t can be expressed using a correlation matrix with N columns. Such correlation matrix named R_x is presented in equation E.5 [208].

$$\hat{\mathcal{R}}_x = \frac{1}{N} \sum_{t=1}^{t=N} x_t x^H(t)\,, \tag{E.5}$$

Two methods have been developed: the MUtliple SIgnal Classification (i.e. MUSIC) method based on the noise subspace and the Estimation of Signal Parameters via Rotational Invariance Techniques (i.e. Esprit) based on the signal subspace. Music and Esprit share this subspace decomposition method as a starting point.

These two methods have been studied in many works in [214]. Such high resolution methods emanating from this subspace decomposition are more robust than the linear prediction techniques one such as Prony [215]. However, despite their high spectral resolution especially on short time windows, high resolution methods imply high computational complexity compared for example to Fourier transform [215].

Music is a high resolution method to estimate directions of arrival of electromagnetic waves arriving on an antenna array. The principle of Music algorithm is based on finding the eigenvalues to obtain the corresponding eigenvectors. To achieve this, Music algorithm uses a set of vectors of the noise subspace to test the orthogonality. Knowing the noise subspace, a research is done on the directions of arrival of electromagnetic wave, such as their direction vector is orthogonal to the noise subspace. A prior estimation of the correlation matrix, that is Rx, is done from measured data. Then, This matrix is diagonalized. Detection of angle of arrival follow these steps:

- Derive a noise subspace base from a diagonalization of the correlation matrix : first the eigenvectors and eigenvalues of the correlation matrix are estimated. Eigenvalues are ranked according to a decreasing order. Then eigenvectors are identified by association with the largest values (e.g. the top first vectors). Such vectors constitute the signal subspace. The rest of the vectors form the noise subspace.

- Extraction of the angles of arrival of emitted electromagnetic waves: using the Music cost function, a projection of this family of vectors of signal subspace is performed on the noise subspace. Then the values of the direction of arrival of the electromagnetic wave is determined. The MUSIC cost function is expressed as follows (cf. equation E.6 [208]):

$$\mathcal{P}_{MUSIC}(\theta) = \frac{a^*(\theta)a(\theta)}{a^*(\theta)\mathcal{E}_n\mathcal{E}_n^* a(\theta)} \tag{E.6}$$

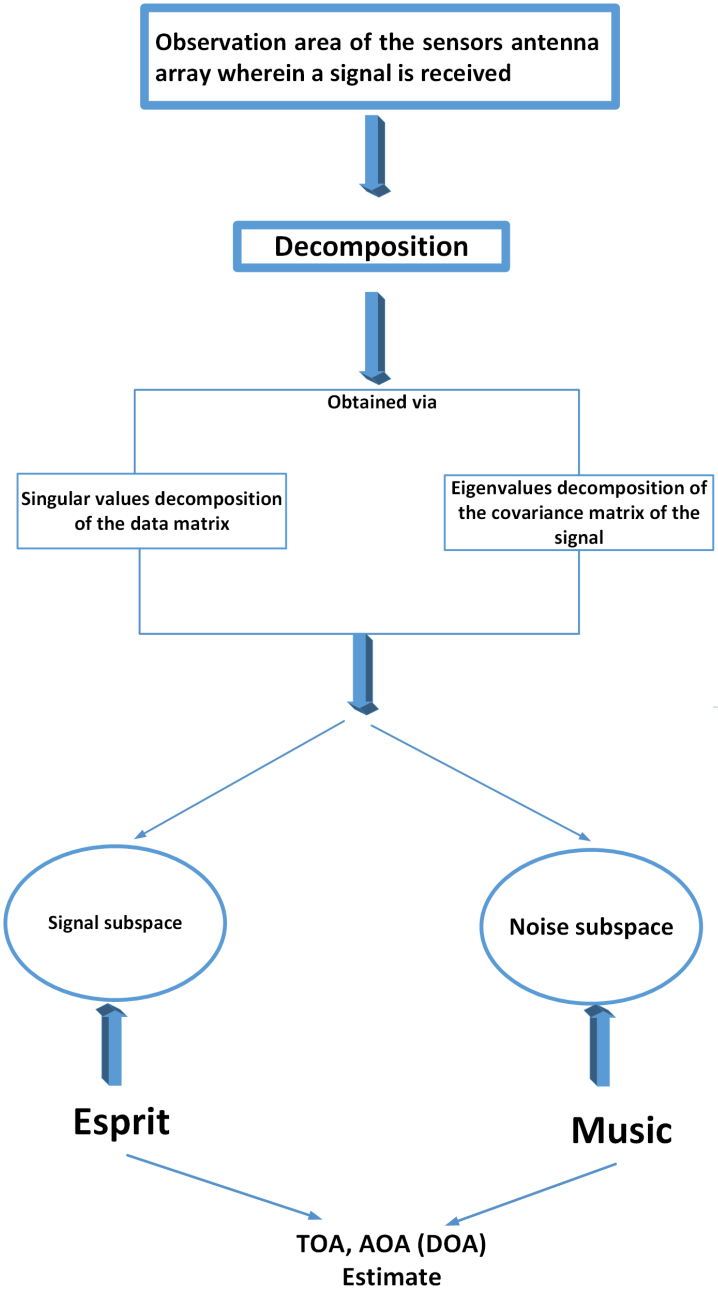


Figure E.2: Overview of the subspace method principle

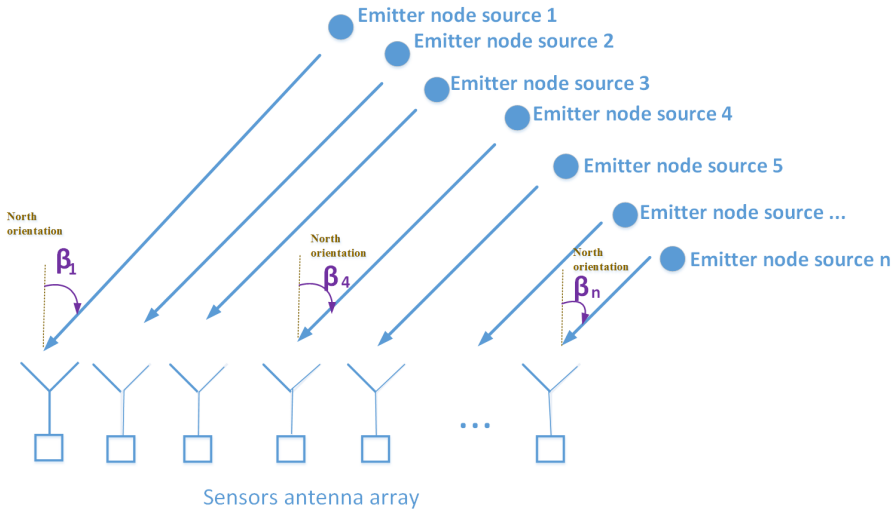


Figure E.3: An antenna array of M sensors nodes and $N = 2$ emitters nodes

Where $a(\theta)$ is the direction vector or impulse response of the sensor array in the direction θ. $E(N)$ is the network of antenna space. Let denote by L the top first vectors that constitute the signal subspace. The condition, $L \leq M - 1$ must be satisfied for the Music algorithm to work. The Music cost function actually looks for which values of θ, corresponding to an angle of arrival, the noise subspace is orthogonal to the signal space. This function is a spectrum depending on θ and maximum for an angle of arrival.

The Music algorithm fits a priori any network structure. Note that it is necessary to have both the signal and noise subspaces for the Music algorithm to work. As mentioned, Esprit is another subspace-based method and it uses the signal subspace. Esprit calculates the directions of arrival (DoA) or angle of arrival of emitted signals (from emitters nodes) relative to a movement of linear antenna arrays. This approach is equivalent to a division of the antenna array into two identical sub antennas networks separated by a vector Δ. Measuring angles relative to the direction of "Delta" is obtained by working with these two sub antenna arrays. The obtained R_x matrix is the sum of both autocorrelations matrices resulting from the two sub antenna arrays. In practice, the array of M sensors are cut into two sub networks of dimension $M - 1$ and Δ will be the gap between the two antennas. Thus, Esprit does not research maxima like Music. Instead, Esprit requires special antennas geometries (e.g. linear and uniform) as the two sub-antennas should be identical [208]. Moreover, a sufficient and necessary condition for the Music and Esprit algorithms to work is that emitters nodes sources must be uncorrelated or must emit uncorrelated signals. This is generally the case when recovering a signal $S(t)$ containing several distinct radio emitters nodes: the signals are uncorrelated in time. This is not the case in the presence of a single emitter node and multiple echo due to multiple paths. Thus, both, Music and Esprit are sensible to multipath.

To pass these limits, improvements have been made to Music and Esprit. Below, we briefly mention the improvements made to Music. But note that the same improvements done with Music were also applied to Esprit.

- The Root version to decorrelate the received signal by doubling the number of measurement samples [216, 217]. A new matrix is obtained and it is the average of two matrices.
- The Smoothing version to increase the order of the autocorrelation matrices [218, 219, 220].

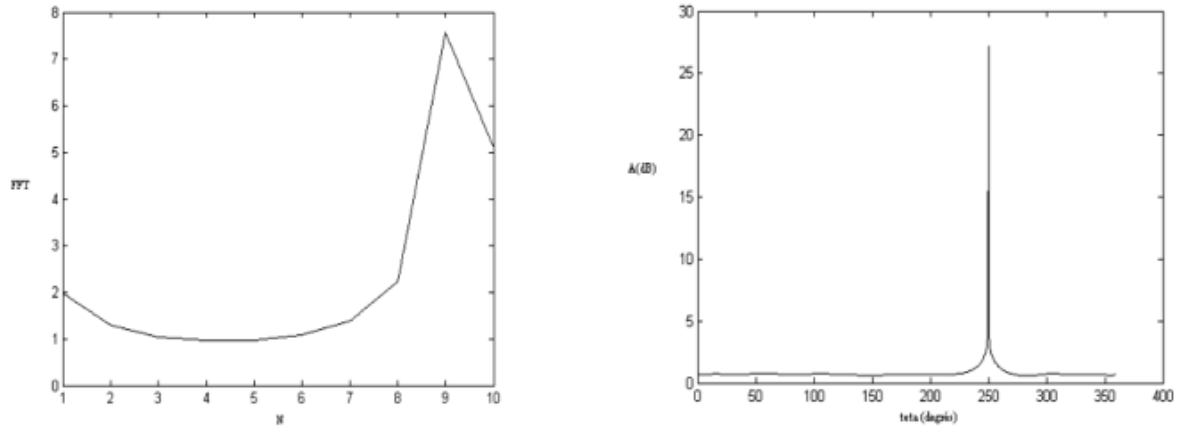


Figure E.4: Shape of the spectra obtained from a linear array of 10 antennas and one emitted node source when using Fourier transform (a) and Music approach (b) [208]

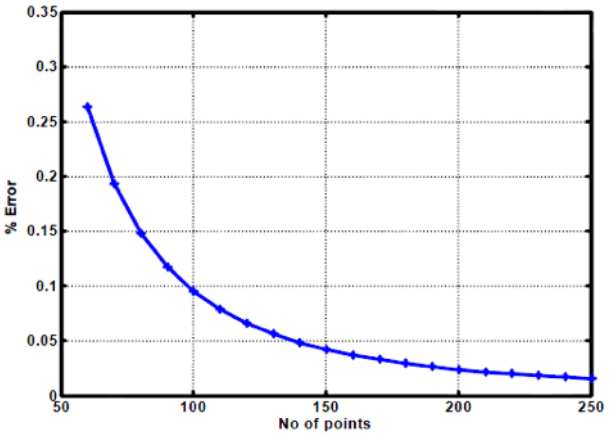
Smoothing Music either decorrelates the correlated emitter nodes whose signal incidence directions differ or the multiple echoes from the same emitter node. Thus, the sensor network let say M is divided into several sub-networks (i.e N). The autocorrelation matrix Rx is computed as the sum of the autocorrelation matrices of the sub antenna arrays. It can detect more AoA compared to Music, under the assumption that if the number of measurement samples is insufficient, or if the sources are correlated, the order of the matrix is too low. This gives a too small number of eigenvalues and thus a limited detection of AoA.

- The cyclic method for the detection of several emitters nodes sources, taking into account the cycle-stationary information contained in the signal [221, 222].
- the Unitary version to accelerate the calculations of finding the angle of arrival of the received signal [223, 95].

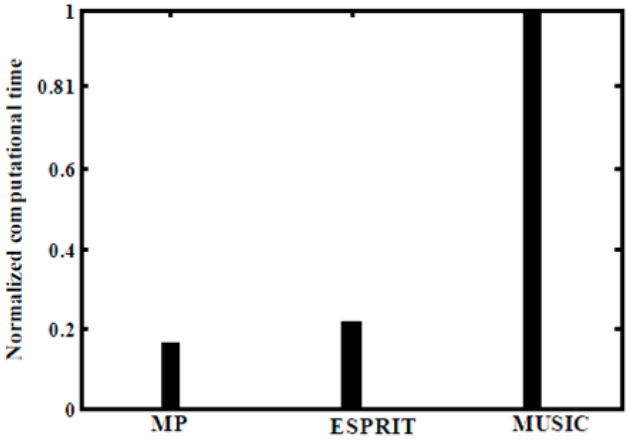
Figure E.4 from the works of De la Roche *et al.* [208] is an example of comparison of two distinct curves obtained with a linear array of 10 antennas from one emitter node source when first using Fourier transform (cf. the left curve) and second applying the Music approach (cf. the right curve).

It is easier to remark that the angle of arrival of the emitter node is accurately and precisely estimated at an angle of 250 ° with the Music approach represented in the right curve in Figure E.4. The angle estimation with Music is also more finer than the Fourier transform. Results in Figure E.4 from the works of De la Roche *et al.* [208] confirm that the Fourier transform is less accurate than the high-resolution methods, for instance, Music.

Another work from Aassie *et al.* [104] presents a ToA estimation using super resolution Matrix Pencil (i.e. MP) in an indoor WLAN positioning system. The matrix pencil algorithm is based on the signal space. As shown in Figure E.5, authors demonstrated in simulation and from experimental results that the Matrix Pencil algorithm has better performance than Esprit when estimating the ToA of a received electromagnetic wave. Esprit algorithm has better computational performance than the root-Music method. The matrix pencil algorithm is also demonstrated to efficiently estimate ToA of a direct line-of-sight path with lower complexity compared to the Inverse Fourier Transform (i.e. IFT). These results can be interpreted with the fact that Music and Esprit are sensible to multipath propagation. It is demonstrated in [104],



(a) Error prediction performance of Matrix Pencil algorithm in ToA estimation in indoor as a function of the number of points of the frequency response for IEEE 801.11 WLAN



(b) Computational time performance of the Matrix Pencil algorithm for a ToA estimation in indoor WLAN compared to ESPRIT and Root-MUSIC

Figure E.5: Comparison of Matrix Pencil, Music and Esprit algorithms when estimating a Time of Arrival of the received signal in an indoor WLAN positioning [104]

to run properly, the MP algorithm needs to identify and to estimate the number of multipath components. Several methods based on statistical classification criteria such as the *Minimum Descriptive Length* (i.e. MDL) and *Akiake information Criterion* (i.e. AIC) [104] are then used to estimate the number of multipath components.

Appendix F

Other localization mechanism based on geometrical techniques

Contents

F.1 Other localization mechanism based on geometrical techniques: Triangulation and Min-Max

F.1.1 Triangulation [1]

Triangulation uses the same mechanism as trilateration, but instead, the identification of the position of an object is based on the angle of arrival estimation of the incoming signal rather than distance. Trilateration involves the distance measurements from at least three landmarks while the triangulation results in angle measurements from at least three landmarks [110]. The multiangulation uses the same principle as multilateration presented in section 4.2.2.2 but with angle measurements from at least four landmarks.

Min-max algorithm or Intersection of quadrilaterals Like trilateration, Min-max is a simple geometric-based localization mechanism based on quadrilaterals intersection that does not require the use of a propagation model to depict the shadowing term from indoor obstacles. This method appeared in [30] (also known as *Min-Max*) as an initialization step to estimate the position of a blind node from known three landmarks nodes positions $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ and distances separation to the landmarks $\{r_1, r_2, r_3\}$. The idea consists in bounding the position into a quadrilateral whose intersection $[x_i - r_i, x_i + r_i] \times [y_i - r_i, y_i + r_i]$ is expressed as the following bounds :

$$[max(x_i - r_i)\ ,\ min(x_i + r_i)] \times [max(y_i - r_i\ ,\ min(y_i + r_i)]$$

Two estimated distances d_j from RSSI measurements from two different landmarks, will therefore lead to either two equations of circles for trilateration method or two quadrilaterals for min-max. And so on, as shown in Figure F.1, three estimated distances from RSSI measurements from three different landmarks lead to the intersection of three quadrilaterals for min-max.

This principle is illustrated in FigureF.1. Taking the estimated distances $\{\hat{r}_1, \hat{r}_2, \hat{r}_3\}$ following the noisy model in equation 4.17 as in the example in Figure F.1, the bias of the estimated values being the centroid of the quadrilateral and resulting from this method is shown in Figure F.2.

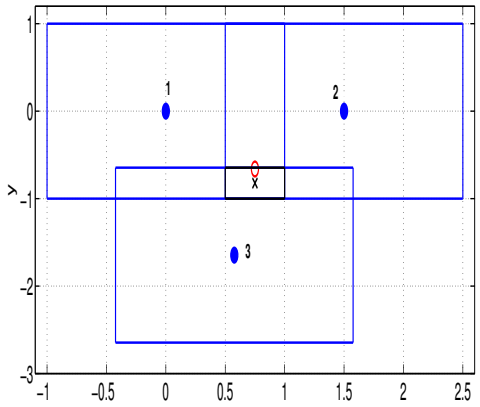


Figure F.1: Intersection (-) of three quadrilaterals (-) when considering values in the example in Figure 4.9 with three equidistant circles ($r = 1$) and therefore $r_1 = r_2 = r_3 = 1$. In this case, the position of the blind node, estimated as the centroid (x) of the quadrilaterals does not coincide with the true position(o) obtained by the expressions in equation 4.16

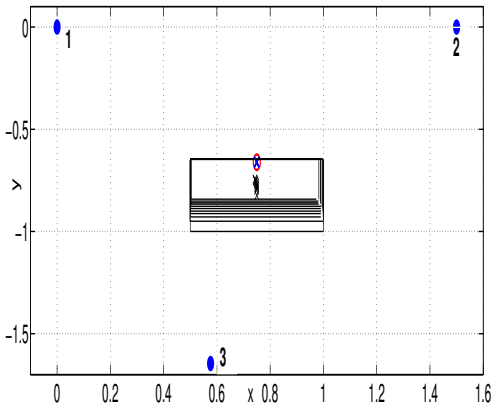


Figure F.2: For different values of σ^2, drawing the average of 10000 Monte Carlo simulation (RSSI data generated with Matlab) which result in $\{\hat{r}_1, \hat{r}_2, \hat{r}_3\}$ of quadrilaterals (-) and the true position (o), estimated by Min-Max (x) and the estimated positions with 4.19 (x)

As noted in [30] and analyzed in detail in [160] and in [6], the Min-Max positioning result is highly dependent on the locations of landmarks that are around the blind node. In addition, with less computational time than localization-based maximum likelihood estimator, min-max does not perform well when the number of landmarks is high [5], like more than 8 landmarks as this may lead to situation with no intersection point. Some configurations assess, for example, if the position to bound is outside the convex envelope formed by the three landmarks, the resulting intersection area in this case is large.

If a good configuration is considered as in the example in Figure F.1 and when calculating the mean of estimated positions with Min-Max (cf. Figure F.2), the component y is always biased by the center of the resulting quadrilaterals which does not coincide with the real position even if $\sigma^2 = 0$ (cf. the behavior of the dashed red line curve in Figure F.3 for small values of σ^2). With a lower noise, the estimated position with Min-Max localization mechanism will always be moved to the center of the quadrilaterals. However, in the case of the presence of lots of noises varying the sides of the quadrilaterals, there has less impact on the centroid of the resulting quadrilateral (as noted by [160], the area of the resulting intersection may sometimes becomes smaller (i.e. in case of destructive interference) and sometimes larger (i.e. in case of constructive

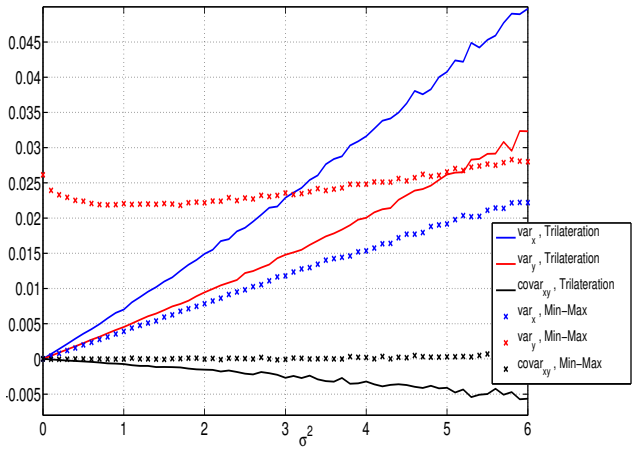


Figure F.3: For different values of σ^2 the normalized mean error (NME) and the covariance obtained from 10000 independent Monte Carlo realizations (simulated RSSI data generated with Matlab) of $\{\hat{r}_1, \hat{r}_2, \hat{r}_3\}$ is plotted. The curves correspond to different values (Normalized mean error of x and of y and the covariance between these two components) with trilateration (-, -, -) and Min-Max (*, *, *).

interference) without no effect on the position of the centroid which does not much vary).

We illustrate the behavior of Min-max in Figure F.3 by comparing the results using the same configuration with trilateration and Min-Max. The black curve shows the components of the covariance of the estimated coordinates with Min-Max being not correlated unlike the case of trilateration. From the curves of the variances (blue and red), Min-Max is less accurate than trilateration in the case of lower noise (i.e. small values of σ^2). But the variance of the estimated position with trilateration increases linearly with the value of σ^2 unlike the behavior of the variance of the position estimated by Min-Max; which tends to be constant even in the case of lots of noises. Thus, Min-Max localization mechanism is more robust against the presence of noise but less accurate than the trilateration algorithm. In addition, the Min-Max localization algorithm running time is 48 seconds which is greater than the trilateration localization algorithm running time: i.e. 8 seconds (cf. Table F.1). Consequently, we implement the trilateration mechanism as a tested one-hop localization mechanism and the one from the geometric methods to compare with our proposed one-hop localization algorithms.

	Algorithm running time (seconds)
Min-Max	48
Trilateration	8

Table F.1: Trilateration and Min-Max running time in seconds when considering 10000 independent Monte Carlo realizations (RSSI data generated with Matlab)

Intersection of Hyperbola This is another geometric-based method that exploits time difference of arrival (TDOA) measurements of two signals received by a blind node [224]. This method aims to exploit the time difference of arrival (TDOA) measurement of two signals received by any sensor node at an unknown position. The hyperbolic equation can be expressed as the difference of two distances between an unknown position $(x_{BN}, y_{BN}, z_{BN})and two known positions$ (x_1, y_1) and (x_2, Y_2) in general:

$\sqrt{(x_{BN} - x_1)^2 + (y_{BN} - y_1)^2} - \sqrt{(x_{BN} - x_2)^2 + (y_{BN} - y_2)^2} = d_1 - d_2$. As in the previous section dealing with the trilateration mechanism, the first landmark (x_1, y_1) is chosen as the

origin of the reference system. This provides the coordinate for four known positions (4.15) in the reference system:

$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$ to $\{(0,0), (d_{12}, 0), (x'_3, y'_3), (x'_4, y'_4)\}$. The three hyperbolic system of equations whose intersection on the point (x'_{BN}, y'_{BN}) can be described as:

$$\begin{aligned}
\sqrt{{x_{BN}}'^2 + {y_{BN}}'^2} - \sqrt{({x_{BN}}' - d_{12})^2 + {y_{BN}}'^2} &= d_1 - d_2 = a \\
\sqrt{{x_{BN}}'^2 + {y_{BN}}'^2} - \sqrt{({x_{BN}}' - x'_3)^2 + ({y_{BN}}' - y'_3)^2} &= d_1 - d_3 = b \\
\sqrt{{x_{BN}}'^2 + {y_{BN}}'^2} - \sqrt{({x_{BN}}' - x'_4)^2 + ({y_{BN}}' - y'_4)^2} &= d_1 - d_4 = c
\end{aligned}$$

Determining the solution (x'_{BN}, y'_{BN}) involves combining the first and the second equation and then using it in the combination of the first and the third equation.

$$\begin{aligned}
A &= \frac{{y'}_4}{c}\frac{b}{{y'}_3}\left(\frac{a}{2} - \frac{b}{2} + \frac{{x'}_3^2 + {y'}_3^2}{2b} - \frac{d_{12}^2}{2a}\right) + \frac{c}{2} - \frac{a}{2} - \frac{{x'}_4^2 + {y'}_4^2}{2c} + \frac{d_{12}^2}{2a} \\
B &= \left(\frac{d_{12}}{a} - \frac{{x'}_4}{c}\right) - \frac{{y'}_4}{c}\frac{b}{{y'}_3}\left(\frac{d_{12}}{a} - \frac{{x'}_3}{b}\right) \\
\text{then :} \quad {x_{BN}}' &= \frac{A}{B} \\
{y_{BN}}' &= \frac{b}{{y'}_3}\left[\left(\frac{d_{12}}{a} - \frac{{x'}_3}{b}\right){x_{BN}}' + \frac{a}{2} - \frac{b}{2} + \frac{{x'}_3^2 + {y'}_3^2}{2b} - \frac{d_{12}^2}{2a}\right]
\end{aligned}$$

Similarly to the previous section on trilateration, the same analysis can be performed. A blind sensor node aims to estimate its relative position (x'_{BN}, y'_{BN}) from measurements of values (a, b, c) with additive noise from at least 3 landmarks nodes. This gives an estimate of $(\hat{x}'_{BN}\hat{y}'_{BN})$ from which the bias and the covariance matrix can be calculated. We can note that the covariance might not be diagonal due to the dependence between the two components $(\hat{x}'_{BN}\hat{y}'_{BN})$.

List of publications

Journal paper

1. N.A. Dieng, C. Chaudet, L. Toutain, T. Ben Meriem, and M. Charbit. "No-calibration localization for indoor wireless sensor networks, ijahuc 2013." *In International Journal of Ad Hoc and Ubiquitous Computing: Special Issue on Localization, Positioning and Coverage in WSNs*, vol. 14, 2013.

Refereed international conference papers

1. N.A. Dieng, C. Chaudet, M. Charbit, L. Toutain, and T. Ben Meriem. "Experiments on the RSSI as a range estimator for indoor localization.," *In New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on*, pages 1–5, 2012.

2. N.A. Dieng, M. Charbit, C. Chaudet, L. Toutain, and T. Ben Meriem, "A multi-path data exclusion model for RSSI-based indoor localization." In *Wireless Personal Multimedia Communications (WPMC), 2012 15th International Symposium on*, pages 336–340, 2012.

3. N.A. Dieng, M. Charbit, C. Chaudet, L. Toutain, and T. Ben Meriem. "Indoor localization in wireless networks based on a two-modes gaussian mixture model." In *IEEE 78th Vehicular Technology Conference (VTC 2013 Fall)*, pages 1–5, Las Vegas, USA, sep 2013.

4. G. Morral, N.A. Dieng, and P. Bianchi. "Distributed On-Line Multidimensional Scaling For Self-Localization In Wireless Sensor Networks." *In 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Pisa, Italy.

5. G. Morral and N.A. Dieng. "Cooperative RSSI-based Indoor Localization : B-MLE and Distributed Stochastic Approximation." *In 2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, pages 1–5, Vancouver, Canada, 14-17 September 2014.

National conference

1. N.A. Dieng, C. Chaudet, L. Toutain, and T. Ben Meriem. "Evaluation and comparison of RSSI based indoor localization algorithms." *In 5e Colloque National sur la Recherche en Informatique et ses Applications (CNRIA'13)*, Ziguinchor, Senegal, apr 2013.

Bibliography

[1] Guoqiang Mao, Baris Fidan, and Brian D.O. Anderson. Wireless sensor network localization techniques. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 51(10), July 2007.

[2] Paramvir Bahl and Venkata N. Padmanabhan. Enhancements to the radar user location and tracking system, technical report, 2000.

[3] J. Lategahn, M. Müller, and C. Röhrig. Tdoa and rss based extended kalman filter for indoor person localization. In *Proceedings of the 78th Vehicular Technology Conference (VTC2013-Fall)*, September 2013.

[4] Sheu C. Shiann-Tsong, Kao Ming-Tse, Hsu Yen-Ming, and Cheng Yen-Chieh. Indoor location estimation using smart antenna system. In *Proceedings of the 78th Vehicular Technology Conference (VTC2013-Fall)*, National Central University, Taiwan, September 2013.

[5] Giovanni Zanca, Francesco Zorzi, Andrea Zanella, and Michele Zorzi. Experimental comparison of rssi-based localization algorithms for indoor wireless sensor networks. In *Workshop on Real-world wireless sensor networks (REALWSN '08)*, Glasgow, Scotland, March 2008.

[6] E. Goldoni, A. Savioli, M. Risi, and P. Gamba. Experimental analysis of rssi-based indoor localization with ieee 802.15.4. In *Wireless Conference (EW), 2010 European*, pages 71–77, 2010.

[7] Karel Heurtefeux and Fabrice Valois. Is rssi a good choice for localization in wireless sensor network? In *IEEE 26th International Conference on Advanced Information Networking and Applications (AINA)*, Fukuoka, Japan, March 2012.

[8] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38:393–422, 2002.

[9] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, January 1992.

[10] Dr. S. C. Liew. Electromagnetic waves, 2001.

[11] Shahin Farahani. *ZigBee Wireless Networks and Transceivers*. Newnes, Newton, MA, USA, 2008.

[12] Nissanka B. Priyantha, Allen K.L. Miu, Hari Balakrishnan, and Seth Teller. The cricket compass for context-aware mobile applications. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, MobiCom '01, pages 1–14, New York, NY, USA, 2001. ACM.

[13] Robert J. Orr and Gregory D. Abowd. The smart floor: a mechanism for natural user identification and tracking. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, pages 275–276, New York, NY, USA, 2000. ACM.

[14] Kurt Partridge, Amstein L, Gaetano Borriello, and T Whitted. Fast intrabody signaling. In *Demonstration at Wireless and Mobile Computer System and Applications (WMCSA)*, dec 2000.

[15] John Krumm, Steve Harris, Brian Meyers, Barry Brumitt, Michael Hale, and Steve Shafer. Multi-camera multi-person tracking for easyliving. In *Proceedings of the Third IEEE International Workshop on Visual Surveillance (VS'2000)*, VS '00, pages 3–, Washington, DC, USA, 2000. IEEE Computer Society.

[16] Jeffrey Hightower and Gaetano Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34(8):57–66, August 2001.

[17] Cesare Alippi, Alan Mottarella, and Giovanni Vanini. A RF map-based localization algorithm for indoor environments. In *IEEE International Symposium on Circuits and Systems (ISCAS 2005)*, Kobe, Japan, May 2005.

[18] D.J. Suroso, P. Cherntanomwong, P. Sooraksa, and J.-i. Takada. Fingerprint-based technique for indoor localization in wireless sensor networks using fuzzy c-means clustering algorithm. In *Intelligent Signal Processing and Communications Systems (ISPACS), 2011 International Symposium on*, pages 1–5, 2011.

[19] Paramvir Bahl and Venkata N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *19th Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom 2000)*, Tel Aviv, Israel, January 2000.

[20] Hui Liu, Pat Darabi, Houshang Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(6), November 2007.

[21] A.A.M. Saleh and R.A. Valenzuela. A statistical model for indoor multipath propagation. *Selected Areas in Communications, IEEE Journal on*, 5(2):128–137, 1987.

[22] Scott Y. Seidel and Theodore S. Rappaport. 914 mhz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation*, 40:207–217, 1992.

[23] Ruonan Zhang and Lin Cai. Modeling uwb indoor channel with shadowing processes. In *4th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*. ACM, 8 2007.

[24] Theodore S. Rappaport. *Wireless Communications: Principles and Practice, Chapter3: mobile radio propagation: large-sacle path loss*. IEEE Press, Piscataway, NJ, USA, 1st edition, 1996.

[25] Theodore S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, second edition edition, 2002.

[26] Yeong-Sheng Chen, Tai-Lin Chin, and Yi-Chen Huang. Collaborative localization in wireless sensor networks based on dependable rssi. In *Wireless Personal Multimedia Communications (WPMC), 2012 15th International Symposium on*, pages 341–347, 2012.

[27] Martin Lauer. A mixture approach to novelty detection using training data with outliers. In *12th European Conference on Machine Learning (EMCL '01)*, London, UK, 2001.

[28] Srdjan Čapkun, Maher Hamdi, and Jean-Pierre Hubaux. Gps-free positioning in mobile ad-hoc networks. *Cluster Computing*, 5(2), April 2002.

[29] Dragos Niculescu and Badri Nath. Ad hoc positioning system (aps). In *IEEE Global Communications Conference (Globecom)*, November 2001.

[30] A. Savvides, H. Park, and M. Srivastava. The Bits and Flops of the N-hop Multilateration Primitive For Node Localization Problems. In *WSNA*, 2002.

[31] Farid Benbadis, Timur Friedman, Marcelo Dias de Amorim, and Serge Fdida. GPS-Free-Free positioning system for sensor networks. In *Second IFIP International Conference on Wireless and Optical Communications Networks (WOCN 2005)*, Dubai, UAE, March 2005.

[32] Pascal Bianchi, Gersende Fort, and Walid Hachem. Performance of a distributed stochastic approximation algorithm. *IEEE Transactions on Information Theory 2013*, abs/1203.1505, Dec 2013.

[33] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized Gossip Algorithms. *IEEE Trans. on Inform. Theory*, 52(6):2508–2530, 2006.

[34] Rajmohan Rajaraman. Antennas and propagation. In *Northeastern University, CS 6710 course*, Spring 2010.

[35] Razvan Musaloiu-E and Andreas Terzis. Minimising the effect of wifi interference in 802.15. 4 wireless sensor networks. *International Journal of Sensor Networks*, 3(1):43–54, 2008.

[36] Chieh-Jan Mike Liang, Nissanka Bodhi Priyantha, Jie Liu, and Andreas Terzis. Surviving wi-fi interference in low power zigbee networks. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, pages 309–322, New York, NY, USA, 2010. ACM.

[37] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, USA, 2005.

[38] Abhijit Mitra. Lecture notes on mobile communication. *A Curriculum Development Cell project Under QIP, IIT Guwahati*, 2009.

[39] Ali Asad. Mobile radio propagation: Large scale path loss, 2011.

[40] Robert Akl, Dinesh Tummala, and Xinrong Li. Indoor propagation modeling at 2.4 ghz for ieee 802.11 networks. In Abraham O. Fapojuwo and Bozena Kaminska, editors, *Sixth IASTED International Multi-Conference on Wireless and Optical Communications: Conference on Communication Systems and Applications, Conference on Optical Communication Systems and Networks, Conference on Wireless Networks and Emerging Technologies,*. IASTED/ACTA Press, 2006.

[41] Fabio Belloni. Fading models s-72.333 physical layer methods in wireless communication systems s-72.333 physical layer methods in wireless communication systems, November 2004.

[42] Lewis Girod. Radio propagation csci 694, March 1999.

[43] Spirent. Testing multipath performance of gnss receivers.

[44] Nitin Jain. An introduction to wireless fading channels, January 2010.

[45] B. Sklar. Rayleigh fading channels in mobile digital communication systems .i. characterization. *Comm. Mag.*, 35(7):90–100, July 1997.

[46] Scott Y. Seidel and Theodore S. Rappaport. A ray tracing technique to predict path loss and delay spread inside buildings. In *Global Telecommunications Conference, . GLOBECOM . IEEE*, 1992.

[47] Christoph Sommer and Falko Dressler. Using the Right Two-Ray Model? A Measurement based Evaluation of PHY Models in VANETs. In *17th ACM International Conference on Mobile Computing and Networking (MobiCom 2011), Poster Session*, Las Vegas, NV, September 2011. ACM.

[48] W. Honcharenko, Henry L. Bertoni, J.L. Dailing, J. Qian, and H. D. Yee. Mechanisms governing uhf propagation on single floors in modern office buildings. *Vehicular Technology, IEEE Transactions on*, 41(4):496–504, 1992.

[49] T.K. Sarkar, Zhong Ji, Kyungjung Kim, A. Medouri, and M. Salazar-Palma. A survey of various propagation models for mobile communication. *Antennas and Propagation Magazine, IEEE*, 45(3):51–82, 2003.

[50] Abdallah Rhattoy and Abdelkarim Zatni. The impact of propagation environment and traffic load on the performance of routing protocols in ad hoc networks. *CoRR*, abs/1202.1677, 2012.

[51] Ali Asad. Mobile radio propagation: Small scale fading and multipath, 2011.

[52] K.L. Du and N.S. Swamy. *Wireless Communication Systems: From RF Subsystems to 4G Enabling Technologies*. Cambridge University Press, 2010.

[53] Theodore S. Rappaport. *Wireless Communications: Principles and Practice, Chapter4: Mobile radio propagation: Smale-scale fading and multipath*. IEEE Press, Piscataway, NJ, USA, 1st edition, 1996.

[54] P. Barsocchi, S. Lenzi, S. Chessa, and G. Giunta. A novel approach to indoor rssi localization by automatic calibration of the wireless propagation model. In *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pages 1–5, 2009.

[55] J. M. Keenan and A. J. Motley. Radio coverage in buildings. *British Telecom Technology Journal*, 8(1):19–24, 1990.

[56] K. W. Cheung, J. H. M. Sau, R. D. Murch, and Hong Kong. A new empirical model for indoor propagation prediction. *IEEE Transactions on Vehicular Technology*, 47(No. 3):pp. 996–1001, 1997.

[57] U. Birkel and M. Weber. Indoor localization with umts compared to wlan. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–6, 2012.

[58] Octavian Modest MANU. A study of indoor localization techniques, 2009.

[59] Veljo Otsason, Alex Varshavsky, Anthony LaMarca, and Eyal de Lara. Accurate gsm indoor localization. In *Proceedings of the 7th international conference on Ubiquitous Computing*, UbiComp'05, pages 141–158, Berlin, Heidelberg, 2005. Springer-Verlag.

[60] Alex Varshavsky, Eyal de Lara, Jeffrey Hightower, Anthony LaMarca, and Veljo Otsason. Gsm indoor localization. *Pervasive Mob. Comput.*, 3(6):698–720, December 2007.

[61] V.J.T. Renaudin-Schouler. *Hybridation MEMS/UWB pour la navigation pédestre intra-muros.* EPFL, 2009.

[62] Zigbee Alliance. Understanding zigbee, http://www.zigbee.org/.

[63] Christian Person. Gt2-3 : Body area network by christian person, page 37, Aug 2011.

[64] Craig Szydlowski. Wireless protocol standards spectrum, May 2004.

[65] Kanji Tanaka, Yoshihiko Kimuro, Kentaro Yamano, Mitsuru Hirayama, Eiji Kondo, and Michihito Matsumoto. A supervised learning approach to robot localization using a short-range rfid sensor. *IEICE - Trans. Inf. Syst.*, E90-D(11):1762–1771, November 2007.

[66] Nosaiba A. Sabto and Khalid Al Mutib. Autonomous mobile robot localization based on rssi measurements using an rfid sensor and neural network bpann. *Journal of King Saud University - Computer and Information Sciences*, 2012.

[67] Hai Liu, M. Bolic, A. Nayak, and I. Stojmenovi. Taxonomy and challenges of the integration of rfid and wireless sensor networks. *Network, IEEE*, 22(6):26–35, 2008.

[68] Jiuqiang Xu, Wei Liu, Fenggao Lang, Yuanyuan Zhang, and Chenglong Wang. Distance measurement model based on rssi in wsn. *Wireless Sensor Network*, 2(8):606–611, 2010.

[69] Raman Kumar K., Varsha Apte, and Yogesh Ashok Powar. Improving the accuracy of wireless lan based location determination systems using kalman filter and multiple observers. In *WCNC*, pages 463–468, 2006.

[70] Tsung-Han Lin, I-Hei Ng, Seng-Yong Lau, Kuang-Ming Chen, and Polly Huang. A microscopic examination of an rssi-signature-based indoor localization system. *The Fifth Workshop on Embedded Networked Sensors (HotEmNets 2008), Virginia, USA*, 2008.

[71] C. Chang. Localization and object-tracking in an ultrawideband sensor network: Research report, 2004.

[72] Jinyun Zhang Jinyun Zhang, P V Orlik, Z Sahinoglu, A F Molisch, and P Kinney. Uwb systems for wireless sensor networks, 2009.

[73] Camillo Gentile and Alfred Kik. A comprehensive evaluation of indoor ranging using ultra-wideband technology. *EURASIP J. Wirel. Commun. Netw.*, 2007(1):12–12, January 2007.

[74] T. Zasowski, F. Althaus, M. Stager, A. Wittneben, and G. Troster. Uwb for noninvasive wireless body area networks: channel measurements and results. In *Ultra Wideband Systems and Technologies, 2003 IEEE Conference on*, pages 285–289, 2003.

[75] Christopher S. Taggart, Yannis Viniotis, and Mihail L. Sichitiu. Modeling the effect of node synchronization times in ultra-wideband wireless networks. *Perform. Eval.*, 66(3-5):223–239, 2009.

[76] Gezici Sinan, Tian Zhi, Biannakis Georgios B., Kobayashi Hisashi, Molisch Andreas F., Poor H.Vincent, and Zafer Sahinoglu. Localization via ultra-wideband radios, July 2005.

[77] Raffaele Bruno and Franca Delmastro. Design and analysis of a bluetooth-based indoor localization system. In Marco Conti, Silvia Giordano, Enrico Gregori, and Stephan Olariu, editors, *PWC*, volume 2775 of *Lecture Notes in Computer Science*, pages 711–725. Springer, 2003.

[78] ficher gunther, DIETRICH Burkhart, and Frank WINKLER. bluetooth indoor localization system. In *PROCEEDINGS OF THE 1st WORKSHOP ON POSITIONING, NAVIGATION AND COMMUNICATION (WPNC'04)*, Taipei, Taiwan, 2004.

[79] University of Missouri-Rolla. What is zigbee? http://web.mst.edu/ mobildat/zigbee/, May 2004.

[80] Pablo Suarez, Carl-Gustav Renmarker, Adam Dunkels, and Thiemo Voigt. Increasing zigbee network lifetime with x-mac. In *Proceedings of the workshop on Real-world wireless sensor networks*, REALWSN '08, pages 26–30, New York, NY, USA, 2008. ACM.

[81] Adam Dunkels. The ContikiMAC Radio Duty Cycling Protocol. Technical Report T2011:13, Swedish Institute of Computer Science, December 2011.

[82] Moteiv Corporation. Chipcon cc2420 datasheet: http://focus.ti.com/lit/ds/symlink/cc2420.pdf, texas instruments, 2007.

[83] Farid Benbadis, Jean-Jacques Puig, Marcelo Dias de Amorim, Claude Chaudet, Timur Friedman, and David Simplot-Ryl. Jumps: Enhancing hop-count positioning in sensor networks using multiple coordinates. available online at http://arxiv.org/abs/cs/0604105, April 2006.

[84] Chao Wang, Kai Liu, and Nan Xiao. A range free localization algorithm based on restricted-area for wireless sensor networks. In *Proceedings of the 2008 The Third International Multi-Conference on Computing in the Global Information Technology (iccgi 2008)*, ICCGI '08, pages 97–101, Washington, DC, USA, 2008. IEEE Computer Society.

[85] Tian He, Chengdu Huang, Brian M. Blum, John A. Stankovic, and Tarek Abdelzaher. Range-free localization schemes for large scale sensor networks. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, MobiCom '03, pages 81–95, New York, NY, USA, 2003. ACM.

[86] Chong Liu, Tereus Scott, Kui Wu, and Daniel Hoffman. Range free sensor localization with ring overlapping based on comparison of received signal strength indicator. *International Journal of Sensor Networks*, 2(5/6):399–413, July 2007.

[87] Marcello Cinque, Christian Esposito, and Flavio Frattini. Leveraging power of transmission for range-free localization of tiny sensors. In *Proceedings of the 11th international conference on Web and Wireless Geographical Information Systems*, W2GIS'12, pages 239–256, Berlin, Heidelberg, 2012. Springer-Verlag.

[88] Koen Langendoen and Niels Reijers. Distributed localization in wireless sensor networks: a quantitative comparison. *Computer Networks - Special issue: Wireless sensor networks*, 43(4), November 2003.

[89] Chris Savarese, Jan M. Rabaey, and Koen Langendoen. Robust positioning algorithms for distributed ad-hoc wireless sensor networks. In *Proceedings of the General Track of the annual conference on USENIX Annual Technical Conference*, ATEC '02, pages 317–327, Berkeley, CA, USA, 2002. USENIX Association.

[90] Qiqian Huang and S. Selvakennedy. A range-free localization algorithm for wireless sensor networks. In *Proceedings of the 63rd IEEE Vehicular Technology Conference, VTC Spring 2006, 7-10 May 2006, Melbourne, Australia*, pages 349–353. IEEE, 2006.

[91] Hongyang Chen, Kaoru Sezaki, Ping Deng, and Hing Cheung So. An improved dv-hop localization algorithm with reduced node location error for wireless sensor networks. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, E91-A(8):2232–2236, August 2008.

[92] Jun Zheng and Asghar Dehghani. Range-free localization in wireless sensor networks with neural network ensembles. *Journal of Sensor and Actuator Networks*, 1(3):254–271, 2012.

[93] Juergen Eckert, Félix Jesús Villanueva, Reinhard German, and Falko Dressler. Distributed mass-spring-relaxation for anchor-free self-localization in sensor and actor networks. In Haohong Wang, Jin Li, George N. Rouskas, and Xiaobo Zhou, editors, *ICCCN*, pages 1–8. IEEE, 2011.

[94] Qing Zhang, Chuan Heng Foh, Boon-Chong Seet, and A. C. M. Fong. Location estimation in wireless sensor networks using spring-relaxation technique. *Sensors*, 10(5), 2010.

[95] Chong Liu, Kui Wu, and Tian He. Sensor localization with ring overlapping based on comparison of received signal strength indicator. *2004 IEEE International Conference on Mobile Adhoc and Sensor Systems IEEE Cat No04EX975*, pages 516–518, 2004.

[96] Flavio Frattini, Christian Esposito, and Stefano Russo. Rocrssi++: An efficient localization algorithm for wireless sensor networks. *Int. J. Adapt. Resilient Auton. Syst.*, 2(2):51–70, April 2011.

[97] Vinay Seshadri, Gergely V Zaruba, and Manfred Huber. A bayesian sampling approach to in-door localization of wireless devices using received signal strength indication. In *Pervasive Computing and Communications, 2005. PerCom 2005. Third IEEE International Conference on*, pages 75–84. IEEE, 2005.

[98] Shinsuke Hara, Dapeng Zhao, Kentaro Yanagihara, Jumpei Taketsugu, Kiyoshi Fukui, Shigeru Fukunaga, and Ken-ichi Kitayama. Propagation characteristics of ieee 802.15.4 radio signal and their application for location estimation. In *IEEE Vehicular Technology Conference (VTC Spring)*, Stockholm, Sweden, May 2005.

[99] Yu-Sheng Lu, Chin-Feng Lai, Chia-Cheng Hu, Yueh-Min Huang, and Xiao-Hu Ge. Path loss exponent estimation for indoor wireless sensor positioning. *Transactions on Internet and Information Systems (TIIS)*, 4(3):243–257, 2010.

[100] A Faheem, R Virrankoski, and M Elmusrati. Improving rssi based distance estimation for 802.15.4 wireless sensor networks, 2010.

[101] The Regents of the University of California. Spss software, 2007.

[102] Masashi Sugano, Tomonori Kawazoe, Yoshikazu Ohta, and Masayuki Murata. Indoor localization system using rssi measurement of wireless sensor network based on zigbee standard. In *IASTED Conference on Wireless Sensor Networks (WSN)*, Banff, Canada, July 2006.

[103] Nayef Alsindi, Xinrong Li, and Kaveh Pahlavan. Performance of toa estimation algorithms in different indoor multipath conditions. In *IEEE Wireless Communications and Networking Conference (WCNC)*, Atlanta, GE, USA, March 2004.

[104] Ali. Aassie and A. S. Omar. Time of arrival estimation for wlan indoor positioning systems using matrix pencil super resolution algorithm. *the 2nd Workshop on Positioning*, 2005.

[105] LILLE) F. Boukour (INRETS, Université de Valenciennes) A. Rivenq (IEMN-DOAE, and LILLE) M.Heddebaut (INRETS, 2009.

[106] Stuart A. Golden and Steve S. Bateman. Sensor measurements for wi-fi location with emphasis on time-of-arrival ranging. *IEEE Transactions on Mobile Computing*, 6(10):1185–1198, October 2007.

[107] R. Exel, G. Gaderer, and P. Loschmidt. Localisation of wireless lan nodes using accurate tdoa measurements. In *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, pages 1–6, 2010.

[108] Jun yong Yoon, Jae-Wan Kim, Won-Hee Lee, and Doo-Seop Eom. A tdoa-based localization using precise time-synchronization. In *Advanced Communication Technology (ICACT), 2012 14th International Conference on*, pages 1266–1271, 2012.

[109] André Günther and Christian Hoene. Measuring round trip times to determine the distance between wlan nodes. In *in Proc. of Networking 2005*, pages 768–779. Springer-Verlag, 2005.

[110] Dragos Niculescu, Badri Nath, Dataman Lab, and Dataman Lab. Ad hoc positioning system (aps) using aoa. In *IEEE INFOCOM*, pages 1734–1743, 2003.

[111] Jian-Guo Wang, A.S. Mohan, and T.A. Aubrey. Angles-of-arrival of multipath signals in indoor environments. In *Vehicular Technology Conference, 1996. Mobile Technology for the Human Race., IEEE 46th*, volume 1, pages 155–159 vol.1, 1996.

[112] P Rong and M L Sichitiu. Angle of arrival localization for wireless sensor networks, 2006.

[113] Charles Wood. The vhf omnidirectional range navigation system, 1999-2008.

[114] Mustapha Boushaba, Abdelhakim Hafid, and Abderrahim Benslimane. High accuracy localization method using aoa in sensor networks. *Comput. Netw.*, 53(18):3076–3088, December 2009.

[115] Bin Xu, Guodong Sun, Ran Yu, and Zheng Yang. High-accuracy tdoa-based localization without time synchronization. *IEEE Transactions on Parallel and Distributed Systems*, 24(8):1567–1576, 2013.

[116] Eiman Elnahrawy, Xiaoyan Li, and Richard P. Martin. The limits of localization using signal strength: a comparative study. In *First Annual IEEE Conference on Sensor and Ad-hoc Communications and Networks*, October 2004.

[117] Chien sheng Chen, Yi jen Chiu, and Jium ming Lin. Hybrid toa/aoa schemes for mobile location in cellular communication systems, June 2010.

[118] Jehn-Ruey Jiang, Chih-Ming Lin, Feng-Yi Lin, and Shing-Tsaan Huang. ALRD: AoA localization with RSSI differences of directional antennas for wireless sensor networks. In *International Conference on Information Society (i-Society)*, Guelph, Ontario, Canada, June 2012.

[119] Chong Liu, Kui Wu, and Tian He. Sensor localization with ring overlapping based on comparison of received signal strength indicator. In *IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS)*, Fort Lauderdale, Florida, USA, October 2004.

[120] Zhang Jianwu and Zhang Lu. Research on distance measurement based on RSSI of ZigBee. In *ISECS International Colloquium on Computing, Communication, Control, and Management*, 2009.

[121] Mohit Saxena, Puneet Gupta, and Bijendra Nath Jain. Experimental analysis of rssi-based location estimation in wireless sensor networks. In *In Proc. Int. Conf. Communication System Software and Middleware*, pages 503–510, 2008.

[122] Wen-Hsing Kuo, Yun-Shen Chen, Gwei-Tai Jen, and Tai-Wei Lu. An intelligent positioning approach: Rssi-based indoor and outdoor localization scheme in zigbee networks. In *ICMLC*, pages 2754–2759. IEEE, 2010.

[123] Abdalkarim Awad, Thorsten Frunzke, and Falko Dressler. Adaptive distance estimation and localization in wsn using rssi measures. In *Proceedings of the 10th Euromicro Conference on Digital System Design Architectures, Methods and Tools*, DSD '07, pages 471–478, Washington, DC, USA, 2007. IEEE Computer Society.

[124] Stefano Tennina, Luigi Pomante, Fabio Graziosi, Marco Di Renzo, Roberto Alesii, and Fortunato Santucci. Distributed localization, tracking, and automatic personal identification: a solution based on a wireless biometric badge. In *Proceedings of the 4th ACM international workshop on Experimental evaluation and characterization*, WINTECH '09, pages 97–98, New York, NY, USA, 2009. ACM.

[125] Jose A. Costa, Patwari N., and O.Hero A. Distributed Weighted-Multidimensional Scaling for Node Localization in Sensor Networks. *ACM TRANSACTIONS ON SENSOR NETWORKS*, 2:39–64, 2005.

[126] Ali Moshfegh. Localization in wireless sensor network, master of science thesis report, 2010.

[127] Arne Schmitz and Martin Wenig. The effect of the radio wave propagation model in mobile ad hoc networks. In *Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, MSWiM '06, pages 61–67, New York, NY, USA, 2006. ACM.

[128] K. Pawlikowski, H.-D.J. Jeong, and J.-S.R. Lee. On credibility of simulation studies of telecommunication networks. *Communications Magazine, IEEE*, 40(1):132–139, 2002.

[129] Hojung Cha and Sungwon Yang. An empirical study of antenna characteristics toward rf-based localization for ieee 802.15.4 sensor nodes. In *4th European conference on Wireless sensor networks (EWSN'07)*, Delft, The Netherlands, January 2007.

[130] Guang-Zhong Yang. *Body Sensor Networks*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[131] Moteiv Corporation. Tmote sky datasheet, 2006.

[132] Peter Miles. 14λ ground plane antenna, July 2011.

[133] Dimitrios Lymberopoulos, Quentin Lindsey, and Andreas Savvides. An empirical characterization of radio signal strength variability in 3-d ieee 802.15.4 networks using monopole antennas. In *Proceedings of the Third European conference on Wireless Sensor Networks*, EWSN'06, pages 326–341, Berlin, Heidelberg, 2006. Springer-Verlag.

[134] Tsenka Stoyanova, Fotis Kerasiotis, Aggeliki Prayati, and George Papadopoulos. Evaluation of impact factors on rss accuracy for localization and tracking applications in sensor networks. *Telecommunication Systems*, 42(3):235–248, 2009.

[135] Liping Wang and Huimin She. Explore energy efficiency in wireless sensor networks based on received signal strength indicator. ACCESS graduate course, Wireless sensor network programming: an introduction,Book of abstracts,laboratory of communication Networks, school of Electrical Engineering, Royal Institute of technology (KTH), 2008.

[136] K Benkic, M Malajner, P Planinsic, and Z Cucej. Using rssi value for distance estimation in wireless sensor networks based on zigbee. *2008 15th International Conference on Systems Signals and Image Processing*, pages 303–306, 2008.

[137] Kannan Srinivasan and Philip Levis. Rssi is under appreciated. In *Third Workshop on Embedded Networked Sensors (EmNets)*, Cambridge, MA, USA, may 2006.

[138] Radim Zemek, Dapeng Zhao, Masahiro Takashima, Shinsuke Hara, Kentaro Yanagihara, Kiyoshi Fukui, Shigeru Fukunaga, and Ken-ichi Kitayama. A Traffic Reducing Method for Multiple Targets Localisation in an IEEE 802.15.4 Based Sensor Network. In *IEEE Vehicular Technology Conference (VTC Fall)*, Montreal, Quebec, Canada, September 2006.

[139] Benoit Le Texier, Nicolas Montavont, and Géraldine Texier. Portail d'acc'es 'a la plate-forme rammus, September 2010.

[140] PC-ENGINE. alix2d2 system board, 2002-2013.

[141] MEMSIC Inc. Wireless sensor networks, 2013.

[142] Université de Strasbourg. Description technique de la plateforme senslab, sep 2010.

[143] Gate Jocelyn. Réalisation de tests sur un réseau de capteurs: la plateforme senslab: Rapport de stage 5'eme année, september 2010.

[144] Colin Chaballier. D1.1(a): Senslab node hardware, july 2008.

[145] Senslab. Welcome to senslab, 2013.

[146] Cameron Dean Whitehouse, David Culler, Kristofer Pister, and Second Reader Date. The design of calamari: an ad-hoc localization system for sensor networks. Technical report, EECS UC Berkeley, 2002.

[147] Sungwon Yang and Hojung Cha. An empirical study of antenna characteristics toward rf-based localization for ieee 802.15.4 sensor nodes. In *Proceedings of the 4th European conference on Wireless sensor networks*, EWSN'07, pages 309–324, Berlin, Heidelberg, 2007. Springer-Verlag.

[148] Juergen Graefenstein and Mohamed Essayed Bouzouraa. Robust method for outdoor localization of a mobile robot using received signal strength in low power wireless networks. In *ICRA*, pages 33–38. IEEE, 2008.

[149] Tsenka Stoyanova, Fotis Kerasiotis, Aggeliki Prayati, and George Papadopoulos. Evaluation of impact factors on rss accuracy for localization and tracking applications. In *Proceedings of the 5th ACM international workshop on Mobility management and wireless access (MOBIWAC '07)*, Chania, Greece, October 2007.

[150] Joaquim AR Azevedo and Filipe Edgar Santos. Signal propagation measurements with wireless sensor nodes, 2007.

[151] P. MORAVEK, D. KOMOSNY, M. SIMEK, M. JELINEK, and A. GIRBAU, D.and LAZARO. Signal propagation and distance estimation in wireless sensor networks. *In The 33rd International Conference on Telecommunication and Signal Processing, TSP 2010*, pages 35 – 40, 2010.

[152] J. Cota-Ruiz, J.-G. Rosiles, P. Rivas-Perea, and E. Sifuentes. A distributed localization algorithm for wireless sensor networks based on the solutions of spatially-constrained local problems. *Sensors Journal, IEEE*, 13(6):2181–2191, 2013.

[153] Ndeye Amy Dieng, Claude Chaudet, Maurice Charbit, Laurent Toutain, and Tayeb Ben Meriem. Experiments on the rssi as a range estimator for indoor localization. *Fifth IFIP International Conference on New Technologies, Mobility and Security*, May 2012.

[154] James M. Lucas, Michael S. Saccucci, Robert V. Baxley, Jr., William H. Woodall, Hazem D. Maragh, Fedrick W. Faltin, Gerald J. Hahn, William T. Tucker, J. Stuart Hunter, John F. MacGregor, and Thomas J. Harris. Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1):1–29, January 1990.

[155] U.S. NIST agency. Nist/sematech e-handbook of statistical methods : Single exponential smoothing, April 2012.

[156] Patel Alpaben K. and Divecha Jyoti. Modified exponentially weighted moving average (ewma) control chart for an analytical process data. In *Journal of Chemical Engineering and Materials Science*, pages 12–20, Victoria Island, Lagos, jan 2011. Academic Journals.

[157] Keith M. Bower. Using exponentially weighted moving average (ewma) charts, October 2000.

[158] T.O. Oshin, S. Poslad, and A. Ma. Improving the energy-efficiency of gps based location sensing smartphone applications. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, pages 1698–1705, 2012.

[159] Neal Patwari, Alfred O. Hero III, Matt Perkins, Neiyer S. Correal, and Robert J. O'dea. Relative location estimation in wireless sensor networks. *IEEE Transactions on Signal Processing*, 51(8), August 2003.

[160] K. Langendoen and N. Reijers. Distributed Localization in Wireless Sensor Networks: A Quantitative Comparison. *Computer Networks*, 43(4):499–518, 2003.

[161] Abhishek Goswami, Luis E. Ortiz, and Samir R. Das. Wigem: a learning-based approach for indoor localization. In *ACM CoNEXT*, 2011.

[162] Mahsa Honary, Lyudmila Mihaylova, and Costas Xydeas. Practical classification methods for indoor positioning. *The Open Transportation Journal*, 6, 2012.

[163] J.C. Chen, R.E. Hudson, and K. Yao. Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field. *IEEE Trans. on Signal Processing*, 50(8):1843–1854, 2002.

[164] F. Potorti, A. Corucci, P. Nepa, F. Furfari, P. Barsocchi, and A. Buffi. Accuracy limits of in-room localisation using rssi. In *Antennas and Propagation Society International Symposium, 2009. APSURSI '09. IEEE*, pages 1–4, 2009.

[165] Borg I. and Groenen P. *Modern Multidimensional Scaling: theory and applications*. New York: Springer-Verlag, 2005.

[166] J. Leeuw and W.J. Heiser. Multidimensional Scaling with Restrictions on the Configuration. *In P Krishnaiah (ed.), Multivariate Analysis*, V:501–522, 1980.

[167] Shang Y., Ruml W., and Fromherz M. Localization from Mere Connectivity. In *MobiHoc*, Anapolis, 2003.

[168] Oh S., Montanari A., and Karbasi A. Sensor Network Localization from Local Connectivity : Performance Analysis for the MDS-MAP Algorithm. In *ITW*, Cairo, 2010.

[169] S.B. Korada, A. Montanari, and S. Oh. Gossip pca. In *ACM SIGMETRICS Performance Evaluation*, 2011.

[170] Morral G., Bianchi P., and Jakubowicz J. Asynchronous Distributed Principal Component Analysis Using Stochastic Approximation. In *CDC*, Hawaii, 2012.

[171] B. Delyon. Stochastic approximation with decreasing gain: Convergence and asymptotic theory. *Unpublished Lecture Notes, http://perso.univ-rennes1.fr/bernard.delyon/as_ cours.ps*, 2000.

[172] N. Patwari, R. J'Odea, and W. Yanwei. Relative location in wireless networks. In *VTC*, 2001.

[173] Frank Wood. Regression estimation - least squares and maximum likelihood, 2011.

[174] Umrigar Cyrus. Monte carlo methods, 2010.

[175] Frank Vanheel, Jo Verhaevert, Eric Laermans, Ingrid Moerman, and Piet Demeester. Automated linear regression tools improve rssi wsn localization in multipath indoor environment. *EURASIP J. Wireless Comm. and Networking*, 2011:38, 2011.

[176] Guoqiang Mao, Baris Fidan, Guoqiang Mao, and Baris Fidan. *Localization Algorithms and Strategies for Wireless Sensor Networks.* Information Science Reference - Imprint of: IGI Publishing, Hershey, PA, 2009.

[177] Laurent Dupont. Paramétrage quasi-optimal de l'intersection de deux quadriques : théorie, algorithme et implantation: mémoire de these, 2004.

[178] IC.S.T. Mathématique 1999. Les coniques, 2007.

[179] Neal Patwari, Robert J. O'Dea, and Yanwei Wang. Relative location in wireless networks. In *IEEE Vehicular Technology Conference (VTC Spring)*, Rhodes , Greece, May 2001.

[180] Kamol Kaemarungsi and Prashant Krishnamurthy. Modeling of Indoor Positioning Systems Based on Location Fingerprinting. In *INFOCOM*, 2004.

[181] Luis Felipe M. de Moraes and Bruno Astuto A. Nunes. CalibrationFree WLAN Location System Based on Dynamic Mapping of Signal Strength. In *MobyWac*, 2006.

[182] N. Dieng, C. Chaude, M. Charbi, L. Toutain, and M. Tayeb Ben. Experiments on the RSSI as a Range Estimator for Indoor Localization. In *NTMS*, Istanbul, Turkey, 2012.

[183] Aghasi H., Hashemi M., and B. Hossein Khala. A Source Localization Based on Signal Attenuation and Time Delay Estimation in Sensor Networks. *International Journal of Computer and Electrical Engineering*, 4(3):423–427, 2012.

[184] Malisa Marijan, Wendi Heinzelman, Gaurav Sharma, and Zeljko Ignjatovic. Optimal resource allocation for wireless video sensors with power-rate-distortion model of imager. *Circuits and Systems, Midwest Symposium on*, 0:256–259, 2009.

[185] Wendi B. Heinzelman, Anantha P. Ch, IEEE, Anantha P. Chandrakasan, Member, Hari Balakrishnan, , and Hari Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1:660–670, 2002.

[186] Ndeye Amy Dieng, Maurice Charbit, Claude Chaudet, Laurent Toutain, and Tayeb Ben Meriem. A multi-path data exclusion model for rssi-based indoor localization. In *15th International Symposium on Wireless Personal Multimedia Communications, WPMC*, September 2012.

[187] Guillaume Gravier. Elements of statistical modeling : Part 2 mixture models and the em algorithm. *Irisa technical report*, 2008.

[188] F. Caballero, L. Merino, and A. Ollero. A general gaussian-mixture approach for range-only mapping using multiple hypotheses. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[189] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1977.

[190] Florian Wickelmaier. An introduction to mds, sound quality research unit, 2003.

[191] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, third edition, 1996.

[192] Theodore S. Rappaport. *Wireless Communications: Principles and Practice* . Prentice Hall, 2002.

[193] Yi Shang and Wheeler Ruml. Improved mds-based localization. In *INFOCOM*, 2004.

[194] J. Leeuw. Applications of Convex Analysis to Multidimensional Scaling. *Recent Developments in Statistics*, pages 133–145, 1977.

[195] New york University Computer Science Department. Chapter 9: Monte carlo methods, 2006.

[196] A. Benveniste, M. Metivier, and Priouret P. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, 1987.

[197] D. P. Bertsekas and J. N. Tsitsiklis. Parallel and distributed computation: numerical methods. Technical report, Athena Scientific, First Edition, 1997.

[198] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained Consensus and Optimization in Multi-Agent Networks. *IEEE Trans. on Automatic Control*, 55(4):922–938, 2010.

[199] P. Bianchi, G. Fort, W. Hachem, and J. Jakubowicz. Convergence of a distributed parameter estimator for sensor networks with local averaging of the estimates. In *ICASSP*, Prague, 2011.

[200] Pascal Bianchi and Jérémie Jakubowicz. Distributed stochastic approximation for constrained and unconstrained optimization. *CoRR*, abs/1104.2773, 2011.

[201] Haggstrom Michael. Some log-normal distributions, February 2011.

[202] Photis NOBELIS. Statistique: Lois théoriques usuelles-lois log-normales, 209-2014.

[203] Mehdi Essoloh, Cédric Richard, and Hichem Snoussi. Localisation distribuée dans les réseaux de capteurs sans fil par résolution d'un problème quadratique. In *21¡ Colloque GRETSI, Troyes, FRA, 11-14 septembre 2007*. GRETSI, Groupe dÕEtudes du Traitement du Signal et des Images, 2007.

[204] Xiaofan Jiang, Dima Ryazanov, and Jorge Ortiz. Dynamic spring force localization, 2006.

[205] Nana Adow Dankwa. An evaluation of transmit power levels for node localization on the mica2 sensor node: Electrical engineering senior project, yale university, March 2004.

[206] Rombaldi Jean-Etienne. Les coniques, alg‘ebre chap18, 2009.

[207] Marc-André Delsuc. Le traitement du signal: La transformée de fourier, September 1996.

[208] G. De La Roche. *Estimation bande étroite des angles d'arrivée dd'un signal radio-mobile en environnement Indoor*. École supérieure de Chimie Physique Electronique de Lyon, july 2003.

[209] Jérôme Perrin, Bruno Torrésani, Philippe Fuchs, et al. Une fonction de corrélation localisée pour la mise en correspondance des images stéréoscopiques. *Traitement du signal*, 16(1), 1999.

[210] Coquelin Pierre-Arnaud, Rabarisoa Jaonary, and Henri Maitre. La correlation des images sur les balayages fractales, 2004.

[211] J.F. BERCHER. Tf, dirac, convolution, et tutti quanti, September 2001.

[212] Gilbert Strang. *Introduction to Linear Algebra, 4th Edition: Chapter 6, Eigenvalues and Eigenvectors*. math.mit.edu, 2009.

[213] Tom Denton and Andrew Waldron. Linear algebra in twenty five lectures, March 2012.

[214] S. Marcos. *Les méthodes à haute résolution: Traitement d'antenne et analyse spectrale*. Traité des nouvelles technologies. Série Traitement du signal. Hermès, 1998.

[215] R. Badeau. *Méthodes à haute résolution pour l'estimation et le suivi de sinusoïdes modulées: Application aux signaux de musique*. ENST, E. École nationale supérieure des télécommunications, 2005.

[216] K.T. Wong and M.D. Zoltowski. Root-music-based azimuth-elevation angle-of-arrival estimation with uniformly spaced but arbitrarily oriented velocity hydrophones. *Trans. Sig. Proc.*, 47(12):3250–3260, December 1999.

[217] HK Hwang, Zekeriya Aliyazicioglu, Marshall Grice, and Anatoly Yakovlev. Direction of arrival estimation using a root-music algorithm. In *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, volume 2. Citeseer, 2008.

[218] Tie-Jun Shan, M. Wax, and T. Kailath. On spatial smoothing for direction-of-arrival estimation of coherent signals. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(4):806–811, 1985.

[219] R.T. Williams, S. Prasad, A. K. Mahalanabis, and L.H. Sibul. An improved spatial smoothing technique for bearing estimation in a multipath environment. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(4):425–432, 1988.

[220] Chen Qing and Liu Ruolun. On the explanation of spatial smoothing in music algorithm for coherent sources. In *Information Science and Technology (ICIST), 2011 International Conference on*, pages 699–702, 2011.

[221] P. Charge, Yide Wang, and Joseph Saillard. An extended cyclic music algorithm. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 3, pages III–3025–III–3028, 2002.

[222] Zhang-Meng Liu, Zhi-Tao Huang, and Yi-Yu Zhou. Generalized wideband cyclic music. *EURASIP J. Adv. Sig. Proc.*, 2009, 2009.

[223] Marius Pesavento, Alex B. Gershman, and Martin Haardt. Unitary root-music with a real-valued eigendecomposition: a theoretical and experimental performance study. *IEEE Transactions on Signal Processing*, 48(5):1306–1314, 2000.

[224] T. S. Rappaport, J. H. Reed, and B. D. Woerner. Position location using wireless communications on highways of the future. *IEEE Communications Magazine*, 34(10):33–41, 1996.